NPS-OR-96-008

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Research in Quantitative Bioassay Methodology and Risk Analysis and Characterization

by

Donald P. Gaver
Patricia A. Jacobs

July 1996

Approved for public release; distribution is unlimited.

Prepared for: U.S. Army Medical Research and Material Command
Fort Detrick, MD 21702-5012

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943-5000

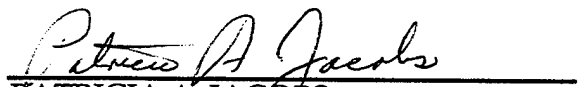
Rear Admiral M. J. Evans
Superintendent

Richard Elster
Provost

This report was prepared for and funded by U.S. Army Medical Research and Materiel Command, Fort Detrick, MD 21702-5012.

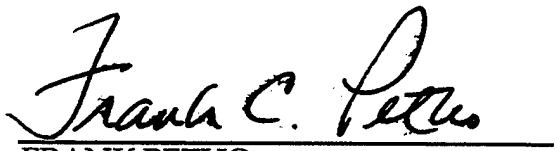
This report was prepared by:

DONALD P. GAVER, JR.
Professor of Operations Research



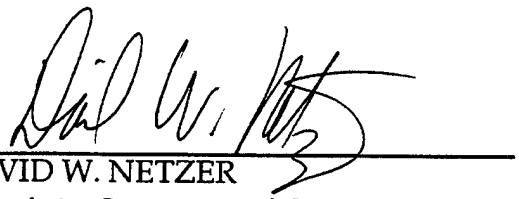
PATRICIA A. JACOBS
Professor of Operations Research

Reviewed by:



FRANK PETHO
Chairman
Department of Operations Research

Released by:



DAVID W. NETZER
Associate Provost and Dean of Research

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1996	Technical

4. TITLE AND SUBTITLE
Research in Quantitative Bioassay Methodology and Risk Analysis and Characterization

5. FUNDING NUMBERS

6. AUTHOR(S)
Donald P. Gaver and Patricia A. Jacobs

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-OR-96-008

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The use of canonical correlation to combine information from biological testing systems is discussed. A graphical procedure to combine results from biological test systems is proposed.

Results are presented of analyses of data from health screens to monitor the health status of medaka used in toxicological studies. A statistical model that incorporates a non-ignorable missing data mechanism is proposed to study the effect of leukocrit values which are not measurable.

Results are presented of analyses of pathology data from the six month interim sacrifice of the West Branch Canal Creek Carcinogenicity Study with Medaka, Test 401-002R.

14. SUBJECT TERMS	15. NUMBER OF PAGES
combining information; multivariate normality; analysis of variance; non-ignorable missing data mechanism; maximum likelihood; logistic regression	120
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

TABLE OF CONTENTS

Research in Quantitative Bioassay Methodology and Risk Analysis and Characterization

by

D. P. Gaver and P. A. Jacobs

ABSTRACT

The use of canonical correlation to combine information from biological testing systems is discussed. A graphical procedure to combine results from biological test systems is proposed.

Results are presented of analyses of data from health screens to monitor the health status of medaka used in toxicological studies. A statistical model that incorporates a non-ignorable missing data mechanism is proposed to study the effect of leukocrit values which are not measurable.

Results are presented of analyses of pathology data from the six month interim sacrifice of the West Branch Canal Creek Carcinogenicity Study with Medaka, Test 401-002R.

Subject Terms: combining information; multivariate normality; analysis of variance; non-ignorable missing data mechanism; maximum likelihood; logistic regression

I. INTRODUCTION AND BACKGROUND

The objectives of the above project were formulated in discussion with Mr. Henry Gardner of U.S. Army Biomedical Research and Development Laboratory, Ft. Detrick, Maryland. The project purpose and workscope was stated in the proposal as follows: to perform mathematical, statistical and risk-

analytical work in support of the mission of the U.S. Army Biomedical Research and Development Laboratory (USABRDL).

II. APPROACHES TAKEN AND PROGRESS

We have analyzed data obtained from other researchers supported by USABRDL. People from whom we have received data include

Dr. Marilyn G. Wolfe, Experimental Pathology Laboratories,

Ms. E. Maxine Boncavage-Hennessey, GEO-CENTERS, INC.,

Dr. Donald C. Malins, Pacific Northwest Research Foundation,

Dr. Lorraine E. Twerdok, GEO-CENTERS, INC.

Mr. Thomas Shedd, USABRDL

The results of these analyses have been reported in the project annual reports for 11 Jan 93 – 11 Jan 94 and 11 Jan 94 – 11 Jan 95 and papers presented at annual research review meetings held in 1993 and 1994.

We have proposed statistical methodology and developed mathematical models in response to the following individuals.

Mr. Henry Gardner, USABRDL

Mr. Robert Finch, USABRDL

Mr. David E. Lovelady, GEO-CENTERS, INC.

Dr. Judith Zelikoff, New York University Medical Center,

Dr. James G. Burkhart, National Institute of Environmental Health Sciences.

Dr. Lorraine E. Twerdok, GEO-CENTERS, INC.

The resulting models and statistical methodology have been reported in the project annual reports for 11 Jan 93 – 11 Jan 94 and 11 Jan 94 – 11 Jan 95.

In this report research results obtained during the period January 12, 1995 – December 31, 1995 are presented.

During the period January 12, 1995 – December 31, 1995 we have proposed methodology to combine information obtained from various biological testing

systems. During 1994, Dr. L. Twerdok and her colleagues began a series of health screens to establish routine health monitoring in Japanese medaka (Oryzias Latipes) used in toxicological studies. We have analyzed data from medaka health screens 2 through 4 obtained from Dr. L. Twerdok in April 1995. At the request of T. Shedd we have analyzed some data contained in a draft copy of the pathology report of the six month interim sacrifice of the U.S. Army Biomedical Research and Development Laboratory Test 401-002R, West Branch Canal Creek Carcinogenicity Study with Medaka. Brief descriptions of our work performed during January 12, 1995 – December 31, 1995 are given below. Details of the work are provided in Appendices.

A. Statistical Approaches for Combining Information from Biological Test Systems for Complex Contaminant Discrimination

A1. Overview

Here is a discussion of what a canonical data analysis does in the context of biological test systems and hazard assessment.

Canonical methods work most directly to compress ("boil down") information from many observational variables on a single biological system to a *score* (or two or three scores) sensitive to general contamination. This was done at Oak Ridge by Adams, Ham, and Beauchamp (Adams, *et al.* (1994)). If a battery of biological test systems is used, as by Burton at Beach Point, this methodology must be extended or replaced *if* an overall score from all systems is wanted. *One way*: summary scores from each system can be derived (for example, the "first canonical variable" score which explains most of the difference between the contaminated site and a reference site using measurements from one biological test system), and a test of the hypothesis of a difference between reference/control score for each system conducted, with result summarized by p-value (small means system sees a difference). These system p-values can be

numerically combined (Fisher's formula, or other, cf. Folks (1984)) or graphed to assess overall evidence for toxicity = hazard at a site.

One approach is to graph the *ordered*, n(= number of systems) p-values vs $1/n+1$: this plot should be 45°-linear if there is no discrimination (details on request).

Note 1: The above techniques do not take account of the *seriousness* in a human or ecological risk sense of the discrimination obtained. If much data is available on one site, and little on another, it may well be that a small (irrelevant) difference between reference and "contaminated" shows up better on the site with the most data. This site may actually be less contaminated and hazardous than the other.

Note 2: The canonical summary isn't the only statistical discrimination tool. We will look into others.

A.2 The Canonical Method

1. Suppose a group of organisms, e.g. medaka, is exposed to a particular complex environment, e.g. suitably buffered full-concentration groundwater from a site for a period of time. Organisms in the group potentially have a *number of responses* to this *dose*. These may be length change, weight change, mobility, leukocrit level, hematocrit level, neoplasms or other organ changes, and other observable features. Some of these are measurable (e.g. length change over dosage period), while others are counted: numbers of fish in a sample at the end of a period exhibiting effect (one or more neoplasms) vs. numbers of fish from a sample at the beginning.

Result: there are *many individual responses* to the above dosage or treatment. These can be coded as a high-dimensional vector of (different) responses.

2. If a comparable group of *the same type of* organisms is exposed to a suitable reference substance, e.g. diluted groundwater, or groundwater from a local

uncontaminated source that is acceptable, then a corresponding (large) set of responses is available for the reference substance.

3. *Problem*: How to treat data in the two data sets so as to efficiently and sensitively quantify the difference between the two sets, for the particular test system, medaka.

4. *Canonical analysis approach*: specifies/derives a *linear* combination (generalized or weighted average) of the *means* of the individual responses, (contaminated and reference) that *best discriminates* between the contaminated group and the reference/uncontaminated group. Alternatively, if you evaluate the linear combination = *score* for each subject (fish) from the contaminated site a cluster of values will occur; likewise for the reference site. The canonical linear combination *separates these clusters* as well as possible. The degree of discrimination is measured by confidence regions around the mean scores, or by the variation of individual scores within clusters. Discrimination is good/effective to the extent that the confidence limits and/or the clusters do not overlap.

Comments

(a) The weights in the above linear combination = score combine the various individual observed responses. It is best when these weights are biologically interpretable. In the Oak Ridge fish study (Adams *et al.* (1994)) they measured 14 variables, but an average of two, namely EROD (enzyme) and BUN (urea nitrogen) explained most of the difference between contaminated and uncontaminated.

(b) In some cases there can be more than one meaningful linear combination = score. A second, or third, such score helps to discriminate further along different (biologically plausible) dimensions. It is best when a very few such scores (*one* is best) does a good job.

(c) The traditional canonical variable technique makes stringent assumptions:

(1) *linear* discrimination is adequate,

(2) normal distributions with *equal* covariance matrices for contaminated and reference responses,

(3) responses are compatible with above; may need to transform, which is possible. More difficult with *counted or categorical* responses.

(d) A Further Problem: there are other biological test systems, e.g. frog embryo, MICROTOX; it is desired to combine data from all of these, suitably weighted.

B. Analysis of Data from Health Screens 2–4 to Establish Routine Health Monitoring in an Aquatic Species (Oryzias Latipes)Used in Toxicological Testing

B1. Introduction

The data consist of measurements made on Japanese medaka (Oryzias Latipes) that were sacrificed at different times during 3 health screens. Health screen 2 occurred during 7/94; health screen 3 occurred during 11/94; and health screen 4 occurred during 1/95.

The information recorded for each fish includes: the date of the experiment (which is called the sacrifice date here); the age (in months); the length (in millimeters); the weight (in milligrams); percent hematocrit; and percent leukocrit. The minimum reported value of leukocrit is 0.01; this value is a code for "unable to measure". There are other missing values which are coded by the value 100. The fish used in the health screens come from several populations. One population consists of fish to be used in immunotox experiments; these fish will be called *experimental*. Another population consists of fish used for breeding; these fish will be called *breeding*; these fish might be stressed due to water

temperature and handling. A third population consists of *retired breeding* stock fish.

Preliminary analyses of data reported in Twerdok *et al.* (1995) and Jacobs and Gaver (1995) suggest that the leukocrit values vary with the experiment date. Twerdok *et al.* indicate that "this variation could result from seasonal variation or be indicative of compromised health status." Further analysis of the leukocrit data is necessary.

B2. Summary of Results Concerning the Ability to Measure Leukocrit and Leukocrit Values

Appendix 1 presents results of analyses of the data to explore the possibility that the ability to measure leukocrit is associated with other covariates. An analysis of variance rejects the null hypothesis that the mean length for fish whose leukocrit values could be measured and the mean length for fish whose leukocrit values could not be measured are equal; (p-value = 0.0002). The mean length for fish whose leukocrit value could not be measured is significantly smaller than the mean length for fish whose leukocrit value could be measured. Thus, it appears that it is more difficult to measure leukocrit in smaller fish. Further, there appears to be a weak association between log leukocrit and length of fish. Shorter fish tend to have higher leukocrit values.

Appendix 1 also reports results from an analysis of the data to investigate possible associations between measured log leukocrit levels and the population (experimental or breeding) the fish are from; the fish whose leukocrit values could not be measured are omitted. The log leukocrit values are used in the analysis to stabilize the variance and symmetrize the leukocrit values since the values are nonnegative and small. Only the age 6 month medaka in health screen 4 exhibit a significant difference (p-value = 9.8×10^{-8}) in the mean log leukocrit

between the two populations. In this case the mean log leukocrit for the breeding population is less than that for the experimental population.

Dr. L. Twerdok asked us to propose statistical methodology to study the leukocrit values that incorporate the information that some leukocrit values are not measurable. She was concerned that those fish for which leukocrit could not be measured might have smaller leukocrit values than those that could. In this event, the fish for which leukocrit values can be measured will give a biased sample of the leukocrit values; their leukocrit values may be larger than usual. This biased sampling effect may provide an explanation for the association of higher leukocrit values with shorter fish. A biased sample of larger than usual leukocrit values may give the mistaken impression that the fish are stressed when they aren't. In an extreme case, unnecessary changes in the procedures used to maintain the medaka would be instituted. If the ability to measure leukocrit is associated with the leukocrit value then the missing data (unable to measure leukocrit) mechanism is said to be non-ignorable. However, the non-ignorable missing data mechanism as presented by the nonmeasurability of leukocrit appears to be little studied; cf. Little and Rubin (1987).

Appendix 2 presents the results of analysis of data from health screens 2–4 to explore possible associations between the ability to measure leukocrit and the value of leukocrit. Exploratory data analysis techniques are used. A formal statistical model is also proposed and the maximum likelihood estimates obtained. The results indicate that the ability to measure leukocrit may be associated with the leukocrit value but the association does not appear to be a large effect. However, analysis of data from additional health screens and biological insight are needed to resolve the issue. The parameter estimates of a model which includes the non-ignorable missing data mechanism still suggest an association between log leukocrit and log weight for the breeding population of

age 6 month medaka and the experimental population of age 8 month medaka; (the estimates of the correlation between log leukocrit and log weight are more than two standard deviations away from 0). Medaka with smaller log weights tend to have higher log leukocrit levels. Thus, a model that includes the effect of nonmeasurability of leukocit, still indicates an association between the size of the fish and the value of leukocrit measured. It remains to be determined if this finding is of biological significance.

B3. Summary of Results Concerning Comparison of Experimental and Breeding Populations

Previous analyses (cf. Twerdok *et al.* (1995)) of the data have considered comparisons between populations using one type of measurement at a time (e.g. length). Analyses restricted to one measurement at a time may overlook differences in the association between measurements for different populations.

Appendix 3 describes and applies a standard statistical procedure for comparing vectors of means between two populations. This technique finds the linear combination of the measurements which results in the greatest discrepancy between the two populations; thus it implicitly considers the univariate comparisons and incorporates the variance-covariance matrix of the measurements. The linear combination which results in the greatest discrepancy may not have obvious interpretation. Hence, if a statistically significant difference is found, further data analysis is needed to determine the reason. Finally, the biological significance of the difference needs to be assessed.

Following is a summary of the results. There is no statistically significant difference between the mean vectors of length, log weight, and log hematocrit between the experimental and breeding populations of medaka that are 8 months of age ($p = 0.49$). There is a significant difference ($p = 0.03$) between the mean vectors of length, log weight, and log hematocrit for the breeding and

experimental populations of medaka that are 6 months of age; there are some smaller log hematocrit values in the breeding population. The mean vectors of log leukocrit and log hematocrit are statistically significantly different ($p = 0.0004$) for the breeding population and experimental population of all medaka that have measured leukocrit values. Members of the breeding population tend to have lower leukocrit levels than the experimental population. It remains to be determined if these differences are of biological significance.

C. Analysis of Some Pathology Data from the Six Month Interim Sacrifice of the West Branch Canal Creek Carcinogenicity Study with Medaka, Test 401-002R.

C1. Introduction

On October 31, 1995, Margaret Toussaint, on behalf of Tom Shedd, sent us a draft copy of the pathology report of the six month interim sacrifice of the U.S. Army Biomedical Research and Development Laboratory Test 401-002R, West Branch Canal Creek Carcinogenicity Study with Medaka.

We quote from the final draft report prepared by Experimental Pathology Laboratories, Inc. (1995), hereafter referred to as EPL (1995). In the test, "groundwater was pumped from a well on-site into two flow-through diluter systems in a biomonitoring trailer. One system had water from the West Branch of Canal Creek as the dilution water. The dilution water in the second system was dechlorinated tap water. Throughout the study laboratory control medaka were maintained at Fort Detrick in well water. At 13 days of age medaka were either initiated or not initiated with 10 mg/L diethylnitrosamine (DEN) for 48 hours. Exposure to the groundwater began at 16 days of age. At six months into the study approximately 20 medaka from each exposure group were euthanized for evaluation." Further information can be found in Experimental Pathology Laboratories, Inc. (1995).

C2. Summary of Results

Logistic regression is used to study the association between the occurrence of endpoints and other covariates. The endpoints considered are the presence of hepatocellular adenoma, the presence of hepatocellular carcinoma, the presence of basophilic foci, and the presence of eosinophilic foci. The covariates considered are a constant; amount of DEN the fish is exposed to (0 mg/L or 10 mg/L); % groundwater; and indicator variables $I_{\text{Canal Creek}}$, I_{Male}, I_{Lab}; where $I_{\text{Canal Creek}} = 1$ if the diluent water is from Canal Creek and 0 otherwise; I_{Male} equals 1 if the animal is male and 0 otherwise; I_{Lab} equals 1 if the diluent water is lab water and 0 otherwise. An association between a covariate and the presence of an endpoint is considered to be statistically significant if the parameter estimate is greater than 2 standard deviations away from 0. The results are summarized as follows.

1. The fish exposed to DEN have a statistically significant greater probability of exhibiting each endpoint than fish not exposed to DEN.

2. For animals not exposed to DEN, there is no statistical evidence that the occurrence of any of the endpoints is associated with the type of diluent water, the sex of the animal, or the % groundwater.

3. For animals exposed to DEN:

a. there is no statistical evidence that the occurrence of hepatocellular carcinoma is associated with the type of diluent water, the sex of the animal, nor the % groundwater;

b. the probability of an animal having hepatocellular adenoma is greater for those fish in Canal Creek diluent water than for the other diluent waters;

c. the probability of an animal having basophilic foci is decreased if the animal is male and is decreased if the diluent is Ft. Detrick well water;

d. the probability of an animal having eosinophilic foci is increased if the animal is male. It is also increased with an increase in % groundwater.

The endpoints of basophilic foci and eosinophilic foci are categorical; 0 = not present, 1 = minimal, 2 = slight/mild, 3 = moderate, 4 = moderately severe, 5 = severe/high. Further analysis of the data incorporating the categorical nature of some of the endpoints is done. The endpoints considered are the presence or absence of hepatocellular adenoma, the category of basophilic foci, the category of eosinophilic foci, the category of cystic degeneration in the liver, and the category of hyaline material in the glomeruli of the kidney.

The Kruskal-Wallis procedure is used as an exploratory procedure to look for possible associations between endpoints. The Kruskal-Wallis statistic is a nonparametric one-way analysis of variance using ranks rather than the original measurements. Those associations that are statistically significant (p-value < 0.05) are further explored using a contingency table χ^2 test for independence. The results of the contingency table analyses are summarized below.

1. For fish in Canal Creek diluent

a. Those fish exposed to DEN tend to have higher categories of hyaline material in the glomeruli of the kidney (p-value = 0.03), higher categories of basophilic foci (p-value = 0.02), higher categories of eosinophilic foci (p-value = 0.00004), and have greater incidence of hepatocellular adenoma (p-value = 10^{-6}) than those fish not exposed to DEN.

b. Fish that have hepatocellular adenoma tend to have higher categories of hyaline material in glomeruli of the kidney (p-value = 0.00015) and higher categories of cystic degeneration in the liver (p-value = 0.023).

c. Males tend to have higher categories of eosinophilic foci than the females (p-value = 0.04).

d. Females tend to have higher categories of basophilic foci than males (p-value = 0.02).

2. For fish whose diluent is tap water

   a. Fish exposed to DEN tend to have higher categories of basophilic foci than fish not exposed to DEN, (p-value = 0.002).

   b. Fish exposed to DEN tend to have higher categories of eosinophilic foci than fish not exposed to DEN (p-value = 0.0006).

3. Fish exposed to DEN and using Canal Creek water as the diluent tend to have more hepatocellular adenoma than fish exposed to DEN and using tap water as the diluent (p-value = 0.006).

4. Fish using Canal Creek water as the diluent tend to have higher categories of hyaline material in glomeruli of the kidney than fish using tap water as the diluent, (p-value = 0.00004 for fish not exposed to DEN and p-value = 10^{-9} for fish exposed to DEN).

III. CONCLUSIONS

It is important to control experimental conditions so as to minimize unwanted sources of variability such as tank effects. Unless these sources of variability are controlled, or adjusted for, they will tend to dilute the strength of inferred associations between measured variables and treatments.

During 1994 Dr. L. Twerdok and her colleagues initiated health screens to monitor the health status of medaka used in toxicological studies. During the period January 12, 1995 – December 31, 1995 we have analyzed data from health screens 2–4. At the request of Dr. Twerdok, special attention has been paid to the effect of unmeasurable leukocrit values. It was found that the ability to measure leukocrit is associated with the size of the fish either measured by its length or weight. If the ability to measure leukocrit is also associated with the value of the

leukocrit, then the missing data mechanism is said to be non-ignorable. A statistical analysis which incorporates a non-ignorable missing data mechanism still finds association between leukocrit value and weight. It has not been determined if these associations are biologically significant. Further experimentation and data analysis may be required to determine the probable cause for the variation observed.

We have analyzed some pathology data from the six month interim sacrifice of the West Branch Canal Creek Carcinogenicity Study with Medaka, Test 401-002R. The data consist of multiple endpoints. Statistical analysis would be easier if the data were available on a disk rather than in paper format. Statistical models need to be developed to investigate the possibility of associations between the joint occurrence of different endpoints and experimental parameters. Such statistical models would be useful to obtain more information from sources such as the pathology reports.

REFERENCES

Adams, S.M., K.D. Ham, and J.J. Beauchamp. "Application of canonical variate analysis in the evaluation and presentation of multivariate biological response data," *Environmental Toxicology and Chemistry*, **13** (1994) pp. 1673-1685.

Chatfield, C. and A.J. Collins. *Introduction to Multivariate Analysis*. Chapman and Hall, New York, 1980.

Collett, D. *Modelling Binary Data*. Chapman & Hall, New York, 1991.

Experimental Pathology Laboratories, Inc. "U.S. Army Biomedical Research and Development Laboratory, Test 401-002R, EPL Project Number 406-035, West Branch Canal Creek Carcinogenicity Study with Medaka, six month interim sacrifice, pathology report, 3 volumes," Final Draft, September 21, 1995.

Folks, J.L. "Combination of independent tests" in *Handbook of Statistics*, Vol. 4, P.R. Krishnaiah and P.K. Sen, eds., Elsevier Science Publishers, North-Holland, New York, 1984, pp. 113-121.

Gaver, D.P. and P.A. Jacobs. "Modeling and statistical analysis of bioassay data: medaka cell proliferation under DEN and TCE," Naval Postgraduate School Technical Report, NPSOR-94-014, October 1994.

Gaver, D.P. and P.A. Jacobs. "Modeling and statistical analysis of bioassay data: medaka cell proliferation under DEN and TCE," U.S. Army Biomedical Research and Development Laboratory Research Review, September 20-22, 1994. To appear in the compendium of research reviews.

Gaver, D.P. and P.A. Jacobs. "Modeling and statistical analysis of medaka bioassay data (with appendices 1 and 2)" Compendium of the FY1990 and FY1992 Research Reviews for the Research Methods Branch, Technical Report 9404, U.S. Army Biomedical Research and Development Laboratory, Fort Detrick, Frederick, MD 21702-5010, 1994.

Gaver, D.P. and P.A. Jacobs. "Quantitative analysis of bioassay data from different biomonitoring test systems (BTS)," U.S. Army Biomedical Research and Development Laboratory Research Review, October, 1993.

Gaver, D.P. and P.A. Jacobs. Proposal for Research in Quantitative Bioassay Methodology and Risk Analysis and Characterization, Annual Report, 11 Jan 93 – 11 Jan 94, 1994.

Gaver, D.P. and P.A. Jacobs. Proposal for Research in Quantitative Bioassay Methodology and Risk Analysis and Characterization, Annual Report, 11 Jan 94 – 11 Jan 95, 1995.

Gaver, D.P., D. Draper, P.K. Goel, J.B. Greenhouse, L.V. Hedges, C.N. Morris, and C. Waternaux. *Combining Information: Statistical Issues and Opportunities for Research.* American Statistical Association, Alexandria, VA, 1994.

Gaver, D.P., P.A. Jacobs, and R.L. Carpenter. "Note on an Alternative Mechanism for Logistic Growth," Naval Postgraduate School Technical Report, NPSOR-95-013, November 1995.

Gaver, D.P., P.A. Jacobs, R.L. Carpenter, and J.G. Burkhart. "A Mathematical Model for Intra-Cellular Effects of Toxins on DNA Adduction and Repair," Naval Postgraduate School Technical Report, NPSOR-95-012, November 1995.

Gibbons, J.D. *Nonparametric Methods for Quantitative Analysis*, Second Edition. Academic Sciences Press, Inc., Columbus, OH, 1985.

Jacobs, P.A. and D.P. Gaver. Remarks resulting from a preliminary examination of data sent by Dr. L. Twerdok of GEO-CENTERS, INC. at U.S. Army BRDL in October 1994. In Proposal for Research in Quantitative Bioassay

Methodology and Risk Analysis and Characterization, Annual Report, 11 Jan 94 – 11 Jan 95, 1995.

Little, R.J.A. and D.B. Rubin. *Statistical Analysis with Missing Data*. Wiley, New York, 1987.

Twerdok, L.E., M.W. Curry, J.R. Beaman, and J.T. Zelikoff. "Development of routine health monitoring methods for use with an aquatic model (Oryzias Latipes) used in immunotoxicological testing," Poster Presentation at Society of Toxicology 34th Annual Meeting, Baltimore, MD, March 7, 1995.

Routine Health Monitoring in an Aquatic Species Japanese *Medaka* (Oryzias Latipes) Used in Toxicological Testing, I:

Preliminary Examination of Leukocrit Data from Health Screens 2, 3, and 4, Using Data Obtained 4/20/95 from L. Twerdok

by

D. P. Gaver and P. A. Jacobs

1. Introduction

The overall purpose of the experiments conducted and subsequent data analyses reported here is to establish the normal range of physiological parameter values to be used as biological endpoints for risk analysis. As one aspect of such a risk analysis a collection or sample of biological entities, in this case Japanese *medaka* (Oryzias Latipes), might be subjected to various concentrations of substances (e.g. groundwater) sampled from a possibly contaminated site. Observed endpoint values at these concentration levels are then compared to those of controls. Response to dose (e.g. concentration of groundwater) is then measured as an appropriate difference between control animals and those receiving the (non-zero) dose. Natural biological variability of subject animals must be understood in order that the comparative experiment be adequately designed and statistically analyzed.

The data consist of measurements made on Japanese *medaka* (Oryzias Latipes) that were sacrificed at different times during 3 health screens. Health screen 2 occurred during 7/94; health screen 3 occurred during 11/94; and health screen 4 occurred during 1/95.

The information recorded for each fish includes: the date of the experiment (which is called the sacrifice date here); age (in months); length (in millimeters); weight (in milligrams); percent hematocrit; percent leukocrit; and hatch date. The minimum recorded value of leukocrit is 0.01; this value is a code for "unable to measure". There are missing values which are coded by the value 100.

The fish used in the health screening study come from several populations. One population consists of fish to be used in immunotox experiments; these fish are considered normal. Another population consists of fish used for breeding; these fish are considered to be stressed due to water temperature and handling. A third population consists of retired breeding stock fish.

2. Censored Leukocrit Values

In this section we investigate possible associations between the ability to measure leukocrit and other variables. Table 1 displays the number of occurrences of the value 0.01 for leukocrit by health screen month and age. The table suggests that it may be more difficult to measure leukocrit in younger fish.

Figure 1 displays two boxplots of the fish lengths. The boxplot labeled measurable on the x-axis is for those fish whose leukocrit level could be measured. The boxplot labeled not measurable on the x-axis is for those fish whose leukocrit level could not be measured and were assigned leukocrit value 0.01. All fish used in the health screens are included. The ○ represents the mean length. The 2 ×'s at the end of the lines display "adjacent values". They are the

Table 1

Number of 0.01 Values for Leukocrit by Health Screen and Age

Health Screen

	7/94		11/94		1/95	
Age (in Months)	No. of Data n_f Exp. Fish (No. of 0.01 Values)	No. of Data s_f Breeders (No. of 0.01 Values)	No. of Data n_f Exp. Fish (No. of 0.01 Values)	No. of Data s_f Breeders (No. of 0.01 Values)	No. of Data n_f Exp. Fish (No. of 0.01 Values)	No. of Data s_f Breeders (No. of 0.01 Values)
3			7(2)			5(4)*
4			7(0)	7(1)		
5	5(3)*	5(3)*		6(2)	19(5)	
6	4(0)		7(1)	7(0)	20(5)	25(2)
7						
8	5(2)	5(0)	7(1)	7(0)	31(4)	40(8)
9			5(0)			
10						
11						
12					11(0)	
13						
14						
15						
16						
17						
18						
19					13(0)*	

Total Number of Fish = 247

Number of 0.01 values = 43

* indicates an unusual number of 0.01 values for the number of fish examined using a binomial model with number of trials the number of fish examined and probability of a 0.01 value equal to 43/247.

n_f = number of fish in experimental (normal) population
s_f = number of fish in breeding (stressed) population

smallest and largest points within 1.5 interquartile distance of the quartiles. The two boxplots of data suggest that the variability of the fish lengths is similar for those fish whose leukocrit could be measured and those for which it could not.

An analysis of variance rejects the null hypothesis that the two means are equal; (p-value = 0.0002, $F = 15.6$, $df_B = 1$, $df_W = 25$). Thus the leukocrit level appears to be harder to measure in fish of smaller length. Note, however, that the two boxplots do overlap considerably. Hence there is no "smallest length" for fish for which leukocrit could be measured.

The probability of not being able to measure percent leukocrit in a fish as a function of length can be modeled using a logistic regression model. The model is as follows:

$$P\{\text{not being able to measure leukocrit} \mid \text{length of animal}\}$$
$$= 1/[1 + \exp\{\beta_0 + \beta_1 \times \text{length}\}]$$

with estimated coefficients

$$\hat{\beta}_0 = -4.28 \quad (1.56) \qquad \hat{\beta}_1 = 0.22 \quad (0.06)$$

with ( ) the standard errors. We will say an estimate is significantly different from 0 if its absolute value is greater than twice its standard error. Since 2 times the standard error of $\hat{\beta}_1$ is 2(0.06) = 0.12 which is less than $\hat{\beta}_1$ there is a significant effect of length of the fish in the ability to measure leukocrit.

The estimated model is used to compute the estimated probability of not being able to measure leukocrit for each fish. Figure 2 displays two boxplots. The boxplot labeled not measurable on the x-axis is for those fitted probabilities of not being able to measure leukocrit for the fish whose leukocrit values could not be measured. The boxplot labeled measurable is for those fitted probabilities of not being able to measure leukocrit for the fish whose leukocrit values could be measured. The fitted probabilities for the population whose leukocrit could not be measured can be larger than those for the population whose leukocrit value could be measured.

Conclusion. One factor influencing the ability to measure percent leukocrit appears to be the size of the fish.

3. Measured Leukocrit

In this section we investigate possible associations between measured percent leukocrit values and other variables. The fish with 0.01 leukocrit value are not considered. The logarithm of the percent leukocrit values is computed to stabilize the variance and symmetrize the values since the values are nonnegative and small.

Figures 3 – 5 display boxplots of the log percent leukocrit values versus length of fish for the experimental population for health screens 2, 3, and 4. There is no strong evidence of a difference in mean log percent leukocrits associated with length; all ANOVA p-values are larger than 0.05.

Figures 6 – 8 display boxplots of the log percent leukocrit values versus the length of fish for the breeding population for health screens 2, 3, and 4. There is no strong evidence that the mean log percent leukocrit is associated with length; all ANOVA p-values are larger than 0.05.

Since there is no strong evidence for association between measured percent leukocrit values and length, the measured leukocrit values are grouped together for each health screen and fish population. Figures 9 – 11 display boxplots of the log percent leukocrit values by population for each health screen. Note that only health screen 4 has a significant difference between the mean log percent leukocrit values for the breeding population and the experimental population (p-value = 10^{-8}). In this case the log percent leukocrit mean for the breeding population is below that for the experimental population.

Least squares regression is used to further explore associations. The following model was estimated

$$\log \% \text{ leukocrit} = \beta_0 + (\beta_1 \times \text{length}) + (\beta_2 \times \text{status})$$

where $$\text{status} = \begin{cases} 1 & \text{if fish from experimental population} \\ 2 & \text{if fish from breeder population.} \end{cases}$$

The following estimates were obtained

Estimates
(standard error)

β_0	β_1	β_2	
2.11	-0.07	-.031	$R^2 = 0.18$
(0.38)	(0.01)	(0.08)	s.e. = 0.54

We will say that an estimate is significantly different from 0 if its absolute value is greater than twice its standard error. Thus all of the regression parameter estimates are significantly different than 0. Since the estimate of $\beta_2 < 0$ and the experimental (respectively breeder) population is coded as having status 1 (respectively 2), the regression indicates that the breeder population tends to have lower leukocrit values than the experimental population. There is also an indication that longer (e.g. older) fish also tend to have lower leukocrit values.

To further investigate possible associations between measured percent leukocrit levels and the population (experimental or breeding) the fish were selected from, the log percent leukocrit levels for fish of age 6 months and age 8 months for health screens 3 and 4 are examined. Figures 12 – 15 each display 2 boxplots of log percent leukocrit values; one for the experimental population and the other for the breeding population; also displayed are the p-values from analyses of variance. Analysis of variance indicates that only the age 6 month *medaka* in health screen 4 exhibit a significant difference in the mean log leukocrit between the two populations (p-value = 9.8×10^{-8}, $F = 44.1$, $df_B = 1$, $df_W = 36$). In this case the mean log leukocrit for the breeding population is less than that for

the experimental population. There is also the suggestion that the leukocrit levels in fish from the breeding population are more variable.

Table 2 displays the number of fish in each length group as well as the sample mean and sample standard deviation of the log percent leukocrits for both the experimental population and the breeding population for health screens 2 – 4. The mean of log percent leukocrits for the 27 millimeter fish from the breeding population looks suspiciously low in comparison to the other means in the breeding population.

Table 2
Moments of Log Percent Leukocrit by Length of Fish
Experimental Population (Breeding Population)

Log Percent Leukocrit

Length	No. of data points		Mean		Standard Deviation	
18	1	(1)	-0.69	(0.64)	—	(—)
19	1	(2)	0.00	(0.61)	—	(0.11)
22	5	(6)	0.16	(0.48)	0.60	(0.21)
23	8	(3)	0.03	(-0.14)	0.42	(0.53)
24	8	(12)	0.09	(-0.31)	0.30	(0.63)
25	18	(9)	0.03	(-0.56)	0.37	(0.57)
26	13	(19)	0.00	(-0.53)	0.39	(0.80)
27	12	(11)	-0.36	(-1.04)	0.48	(0.47)
28	12	(6)	-0.19	(-0.63)	0.26	(0.40)

Conclusion. There appears to be a weak association between the length of a fish and the measured leukocrit values. There also appears to be a weak association between leukocrit level and population (experimental or breeding) the fish are sampled from. However, this association may also be affected by other factors such as the tank the fish are sampled from.

REFERENCE

IBM Corporation. A Graphical Statistical System (AGSS).

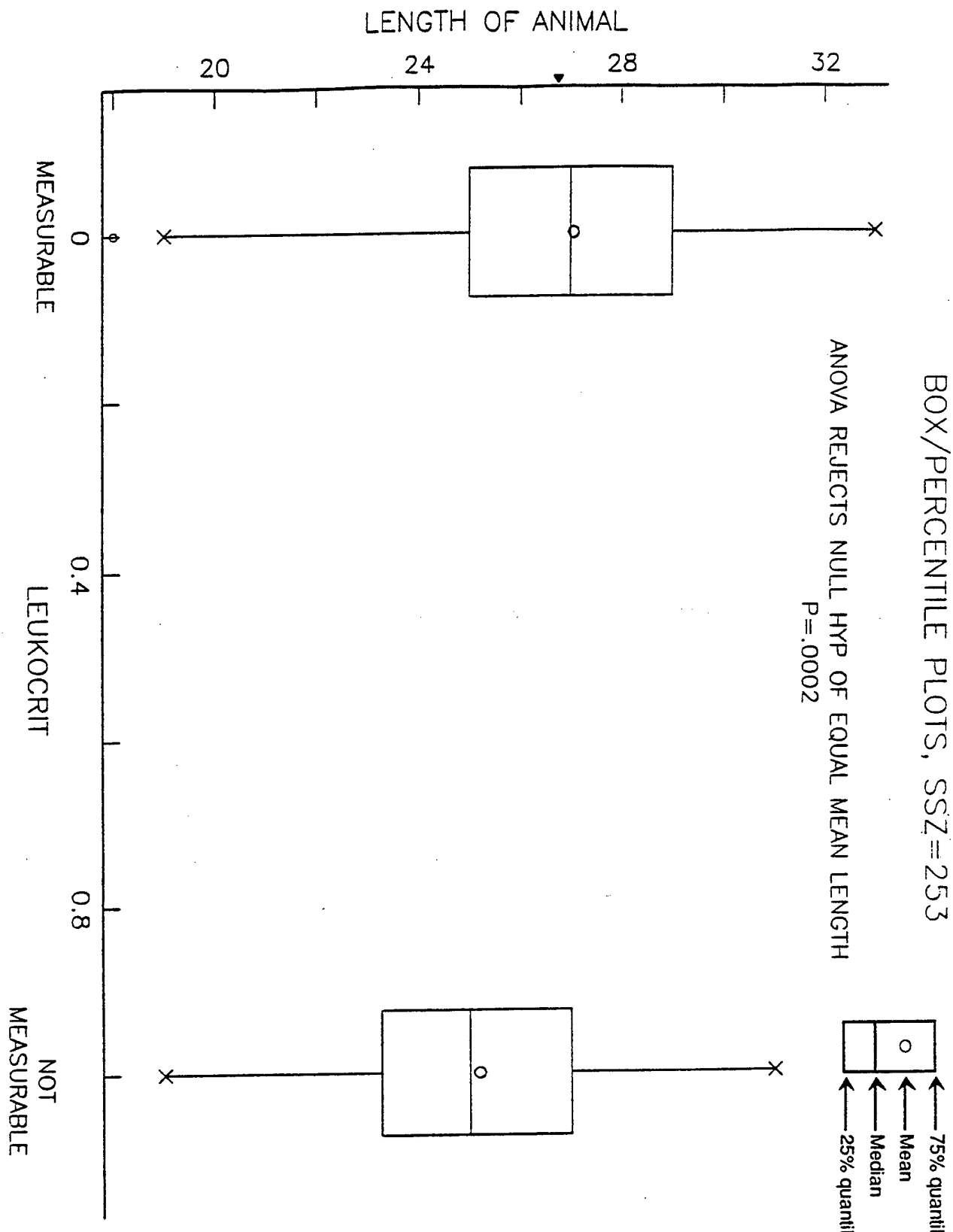

LENGTH OF ANIMAL
20
24
28
32
MEASURABLE
0
0.4
0.8
NOT MEASURABLE
LEUKOCRIT
BOX/PERCENTILE PLOTS, SSZ=253
ANOVA REJECTS NULL HYP OF EQUAL MEAN LENGTH
P=.0002
75% quantile
Mean
Median
25% quantile

Figure 1

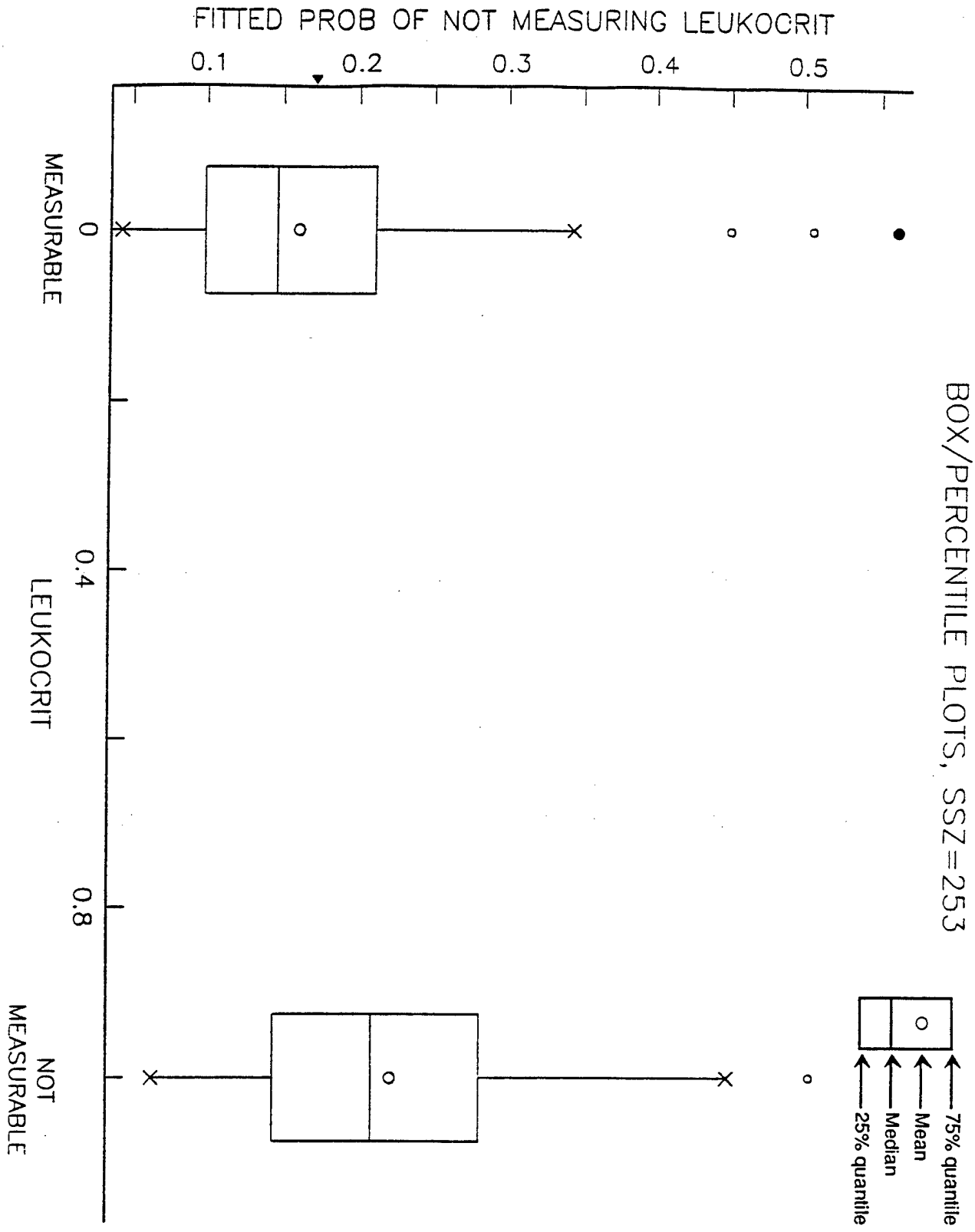

FITTED PROB OF NOT MEASURING LEUKOCRIT
0.1
0.2
0.3
0.4
0.5
MEASURABLE
0
0.4
0.8
LEUKOCRIT
NOT MEASURABLE
BOX/PERCENTILE PLOTS, SSZ=253
75% quantile
Mean
Median
25% quantile

Figure 2

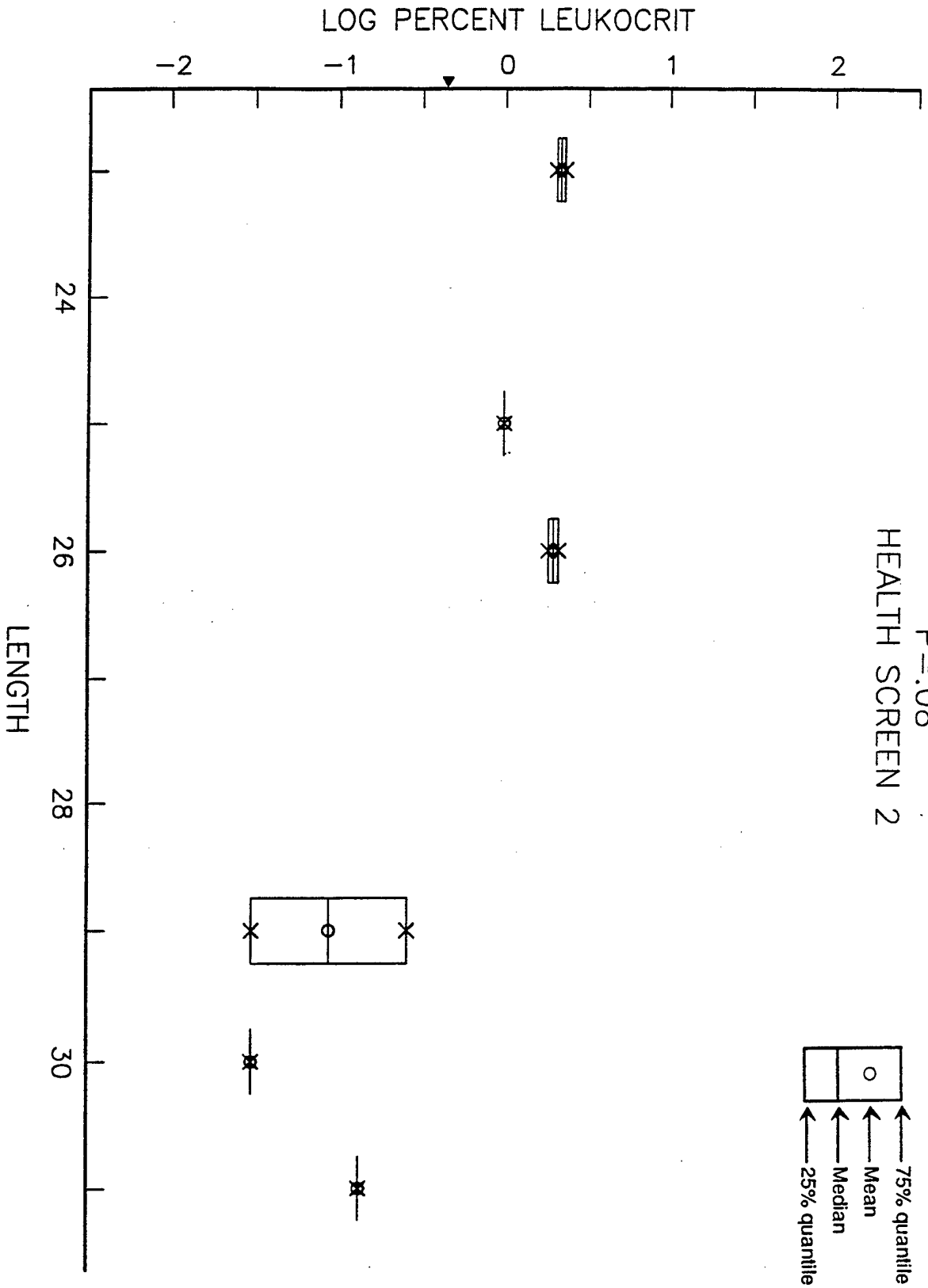

FISH WITH MEASURABLE LEUKOCRIT IN EXPER POPUL
ANOVA DOES NOT REJECT THE NULL HYPOTHESIS OF EQUAL MEANS
P=.08
HEALTH SCREEN 2
LOG PERCENT LEUKOCRIT
−2
−1
0
1
2
LENGTH
24
26
28
30
75% quantile
Mean
Median
25% quantile

Figure 3

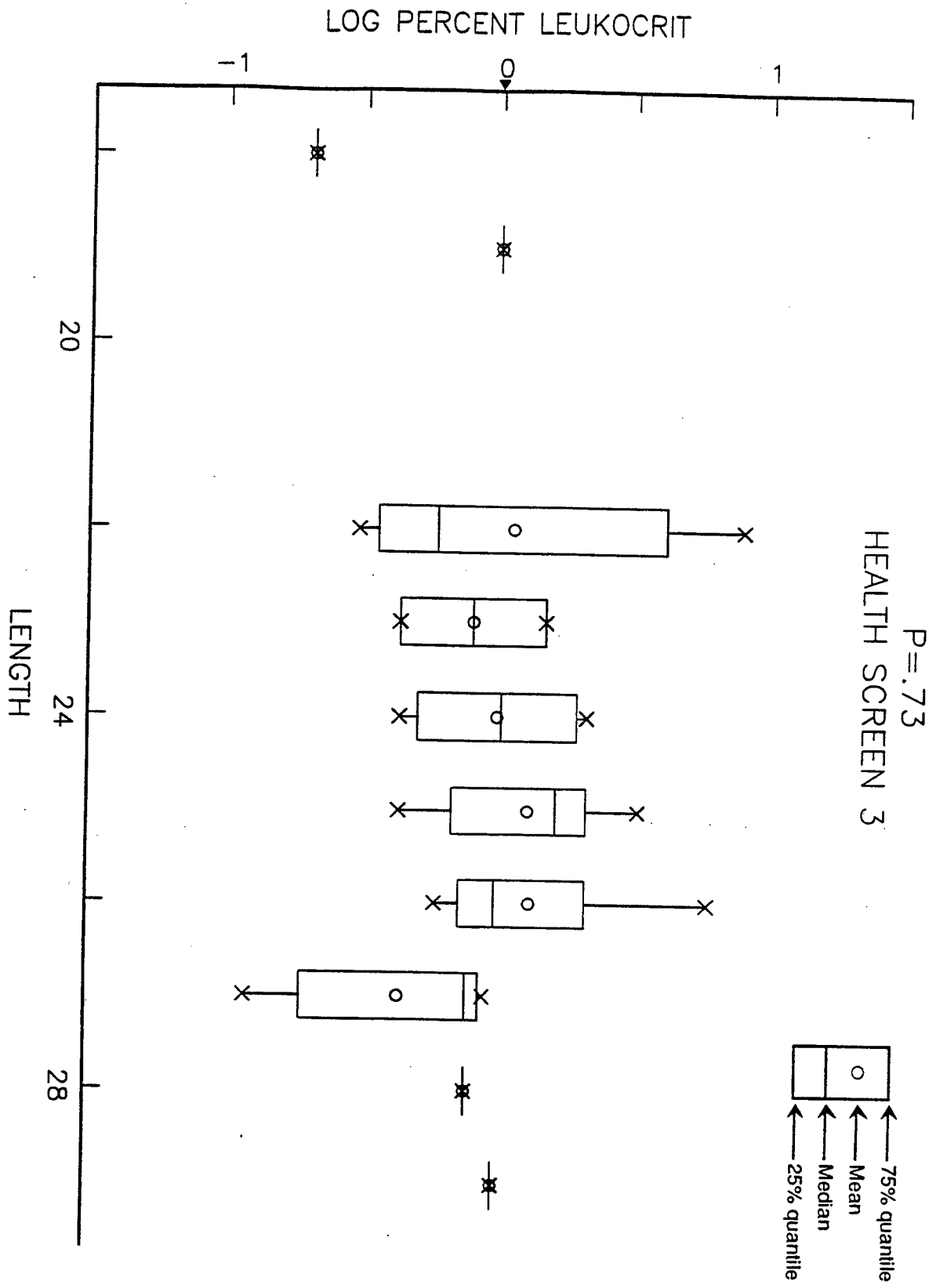

FISH WITH MEASURABLE LEUKOCRIT IN EXPER POPUL
ANOVA DOES NOT REJECT THE NULL HYPOTHESIS OF EQUAL MEANS
P=.73
HEALTH SCREEN 3
LOG PERCENT LEUKOCRIT
−1
0
1
LENGTH
20
24
28
75% quantile
Mean
Median
25% quantile

Figure 4

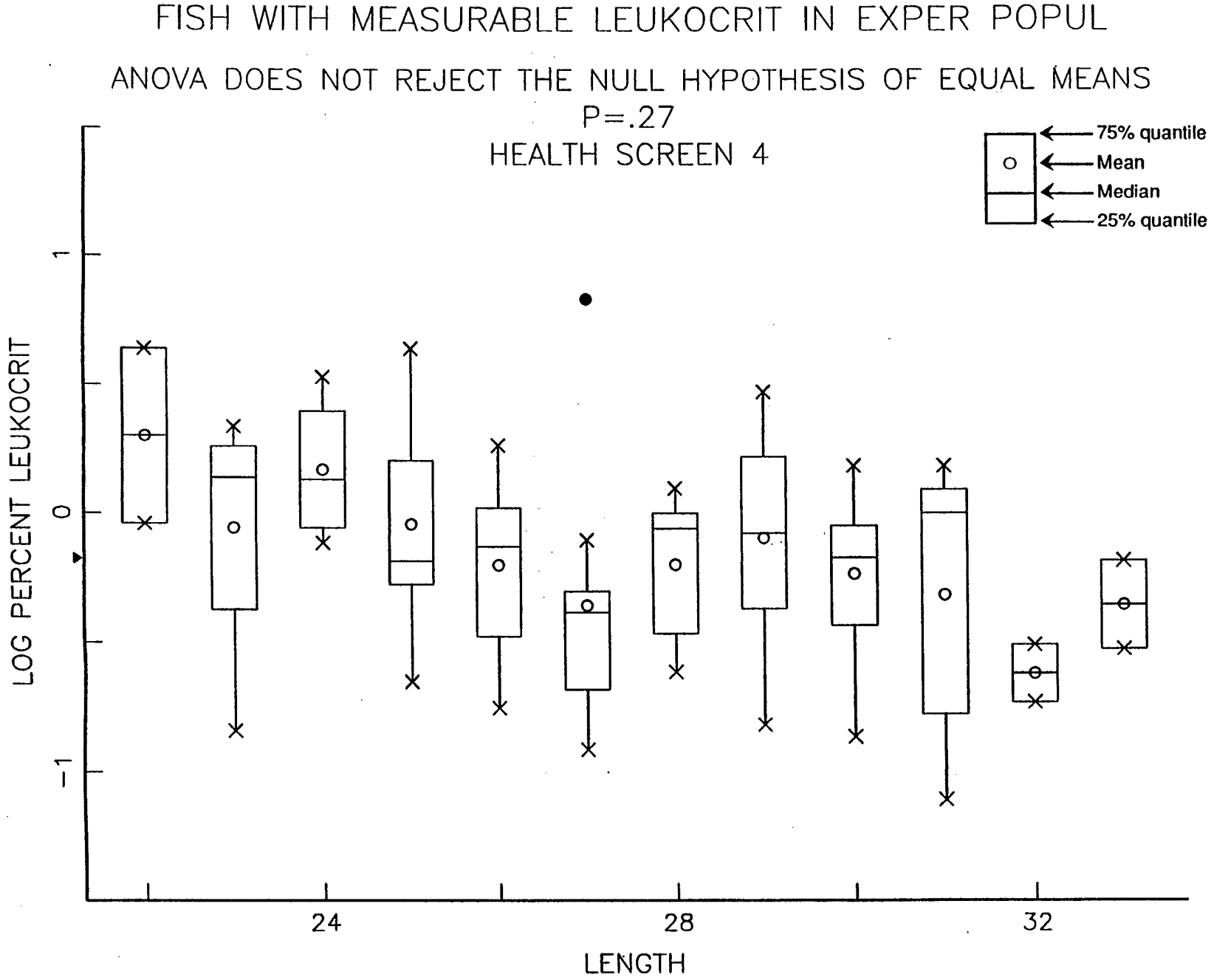

FISH WITH MEASURABLE LEUKOCRIT IN EXPER POPUL
ANOVA DOES NOT REJECT THE NULL HYPOTHESIS OF EQUAL MEANS
P=.27
HEALTH SCREEN 4
75% quantile
Mean
Median
25% quantile
LOG PERCENT LEUKOCRIT
1
0
-1
24
28
32
LENGTH

Figure 5

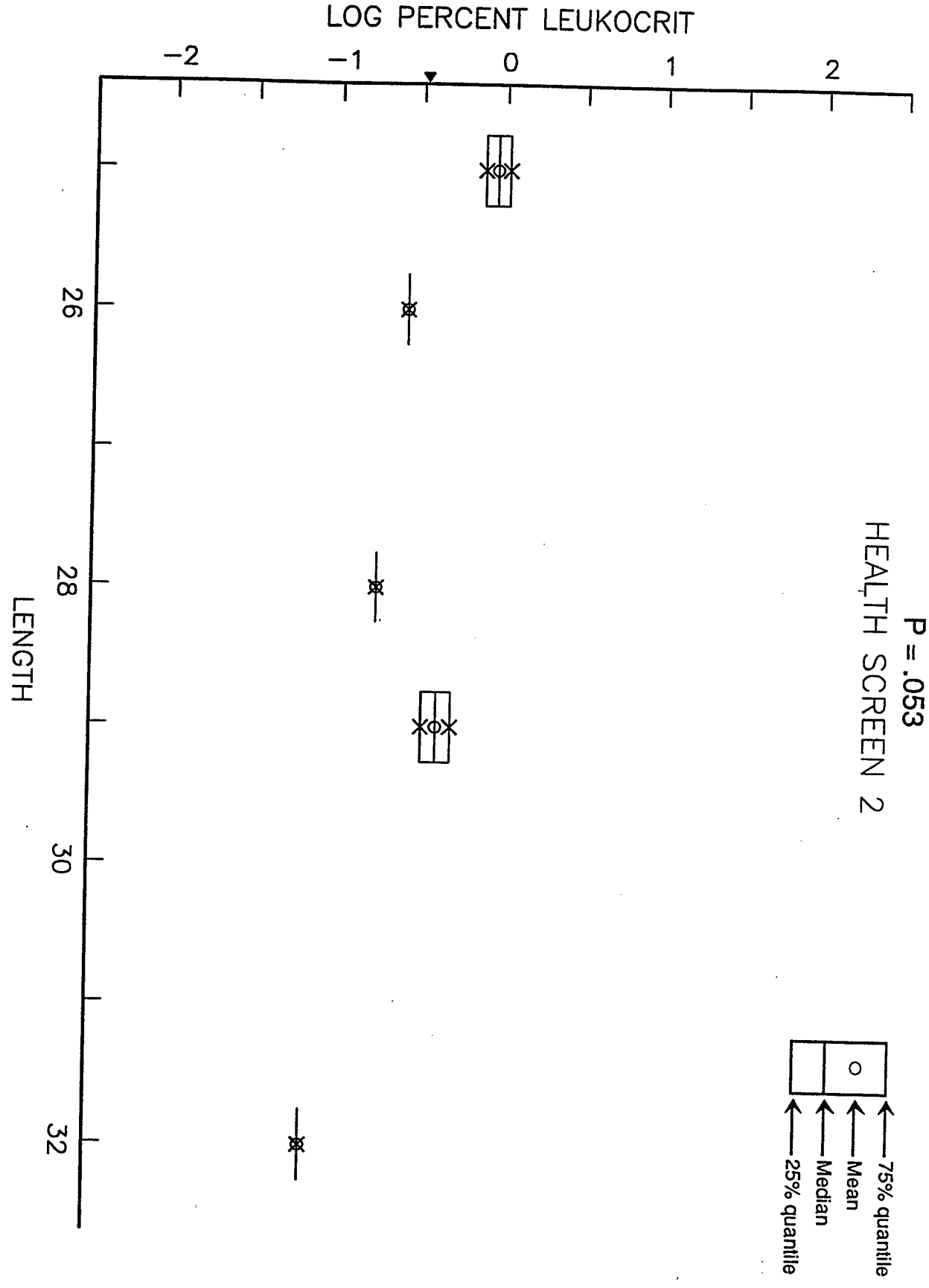

FISH WITH MEAS LEUKOCRIT IN BREED POPULATION
ANOVA DOES NOT REJECT THE NULL HYPOTHESIS OF EQUAL MEANS
P = .053
HEALTH SCREEN 2
LOG PERCENT LEUKOCRIT
−2
−1
0
1
2
LENGTH
26
28
30
32
75% quantile
Mean
Median
25% quantile

Figure 6

FISH WITH MEAS LEUKOCRIT IN BREED POPULATION
ANOVA DOES NOT REJECT THE NULL HYPOTHESIS OF EQUAL MEANS
P=.41
HEALTH SCREEN 3
75% quantile
Mean
Median
25% quantile
LOG PERCENT LEUKOCRIT
2
1
0
−1
−2
20
24
28
32
LENGTH

Figure 7

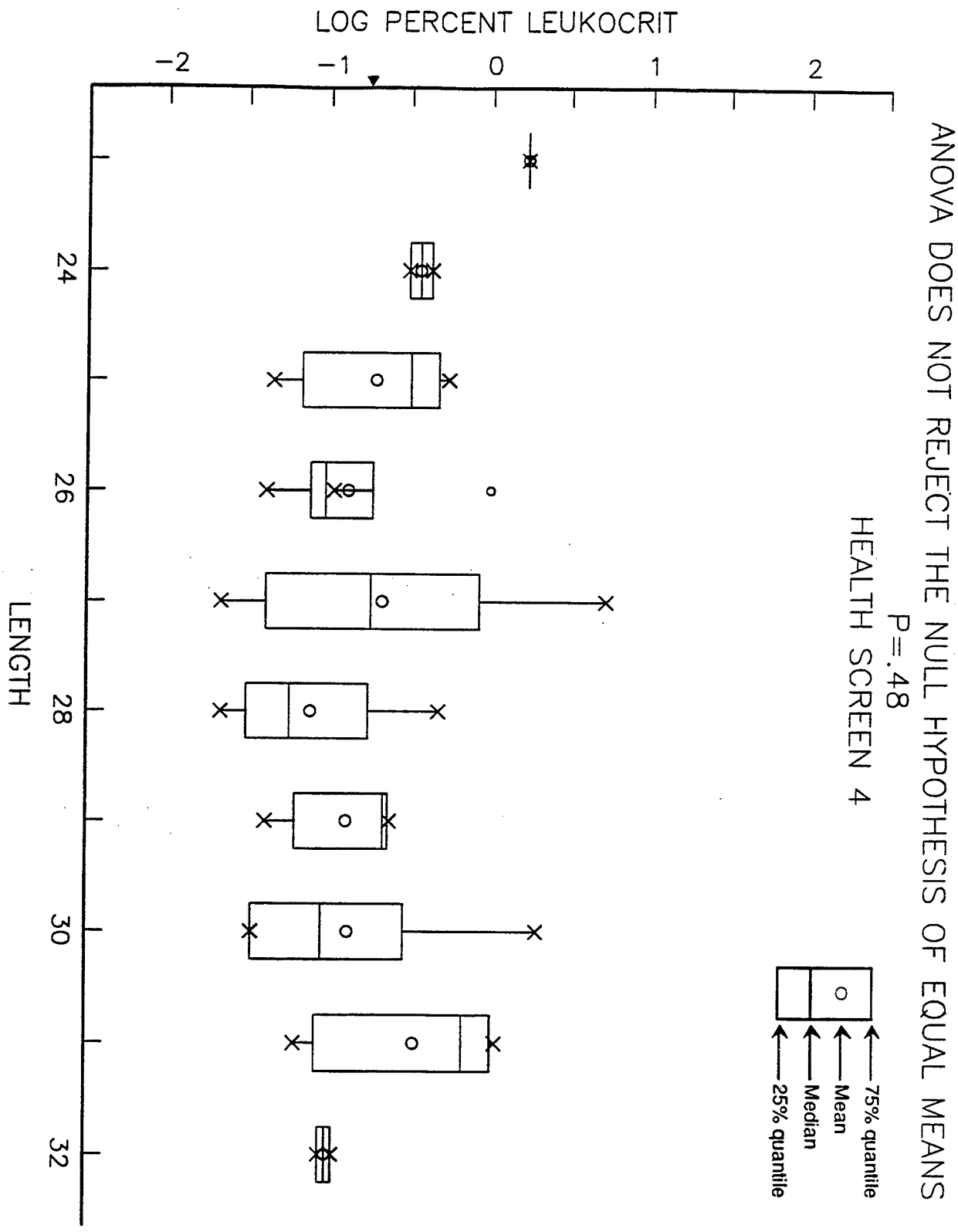

FISH WITH MEAS LEUKOCRIT IN BREED POPULATION
ANOVA DOES NOT REJECT THE NULL HYPOTHESIS OF EQUAL MEANS
P=.48
HEALTH SCREEN 4
LOG PERCENT LEUKOCRIT
−2
−1
0
1
2
LENGTH
24
26
28
30
32
75% quantile
Mean
Median
25% quantile

Figure 8

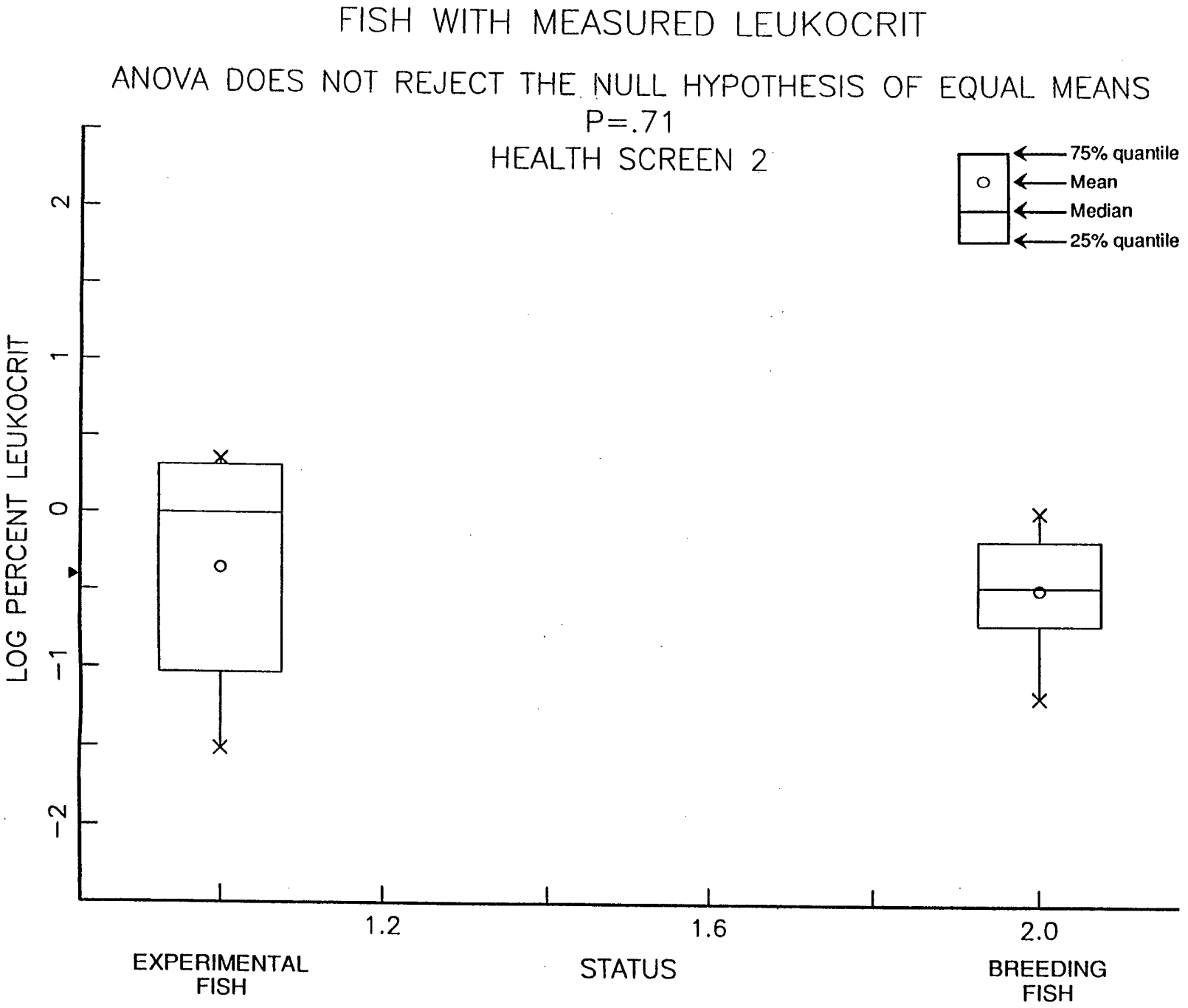

FISH WITH MEASURED LEUKOCRIT
ANOVA DOES NOT REJECT THE NULL HYPOTHESIS OF EQUAL MEANS
P=.71
HEALTH SCREEN 2
75% quantile
Mean
Median
25% quantile
LOG PERCENT LEUKOCRIT
2
1
0
−1
−2
1.2
1.6
2.0
EXPERIMENTAL FISH
STATUS
BREEDING FISH

Figure 9

FISH WITH MEASURED LEUKOCRIT
ANOVA DOES NOT REJECT THE NULL HYPOTHESIS OF EQUAL MEANS
P=.11
HEALTH SCREEN 3
75% quantile
Mean
Median
25% quantile
LOG PERCENT LEUKOCRIT
2
1
0
−1
−2
1.2
1.6
2.0
EXPERIMENTAL FISH
STATUS
BREEDING FISH

Figure 10

FISH WITH MEASURED LEUKOCRIT
ANOVA REJECTS THE NULL HYPOTHESIS OF EQUAL MEANS
P = 1 x (10↑-8)
HEALTH SCREEN 4
75% quantile
Mean
Median
25% quantile
LOG PERCENT LEUKOCRIT
2
1
0
−1
−2
1.2
1.6
2.0
EXPERIMENTAL FISH
STATUS
BREEDING FISH

Figure 11

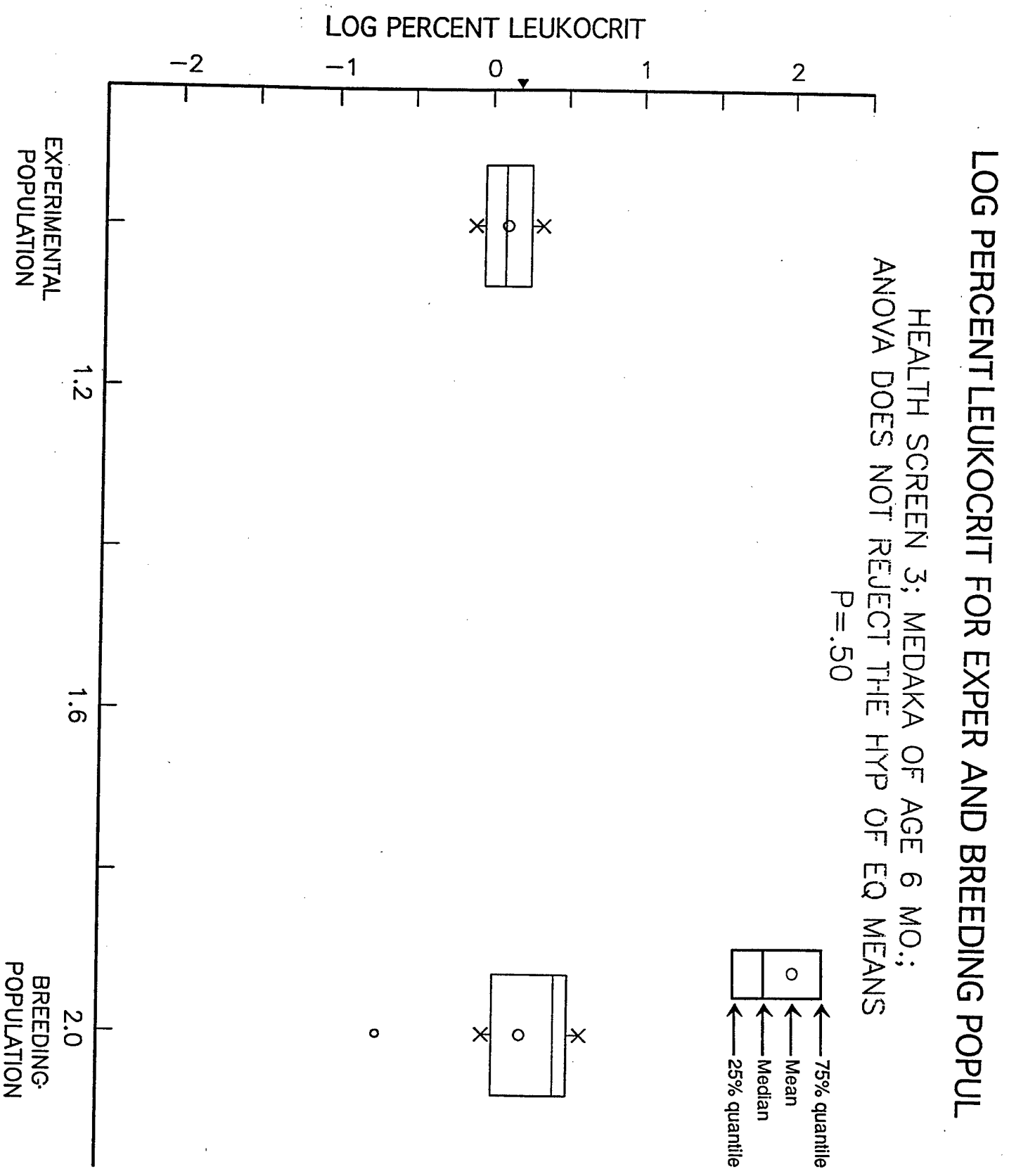

LOG PERCENT LEUKOCRIT FOR EXPER AND BREEDING POPUL
HEALTH SCREEN 3; MEDAKA OF AGE 6 MO.;
ANOVA DOES NOT REJECT THE HYP OF EQ MEANS
P=.50
LOG PERCENT LEUKOCRIT
−2
−1
0
1
2
EXPERIMENTAL POPULATION
BREEDING POPULATION
1.2
1.6
2.0
75% quantile
Mean
Median
25% quantile

Figure 12

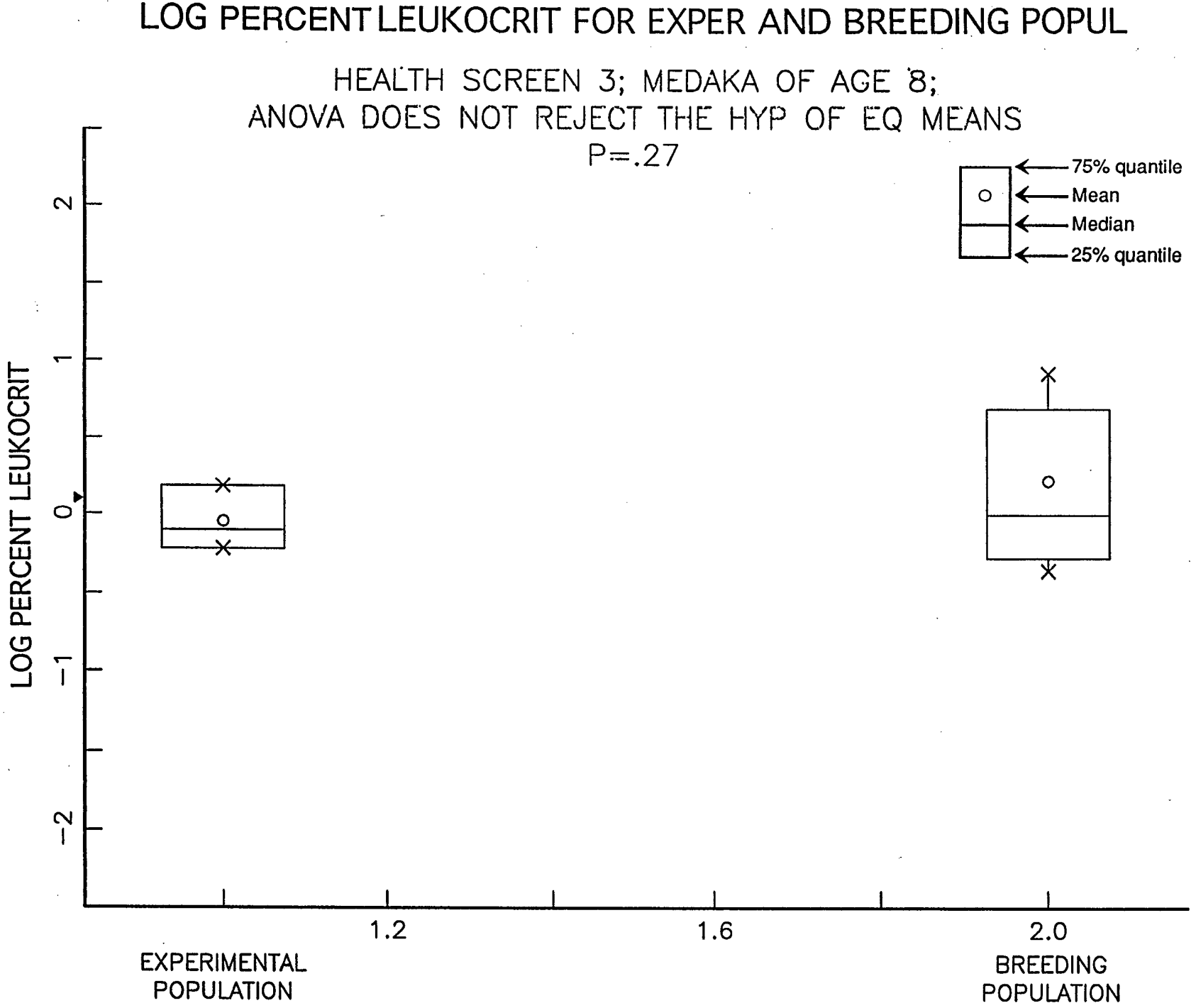

LOG PERCENT LEUKOCRIT FOR EXPER AND BREEDING POPUL
HEALTH SCREEN 3; MEDAKA OF AGE 8;
ANOVA DOES NOT REJECT THE HYP OF EQ MEANS
P=.27
75% quantile
Mean
Median
25% quantile
LOG PERCENT LEUKOCRIT
2
1
0
−1
−2
1.2
1.6
2.0
EXPERIMENTAL POPULATION
BREEDING POPULATION

Figure 13

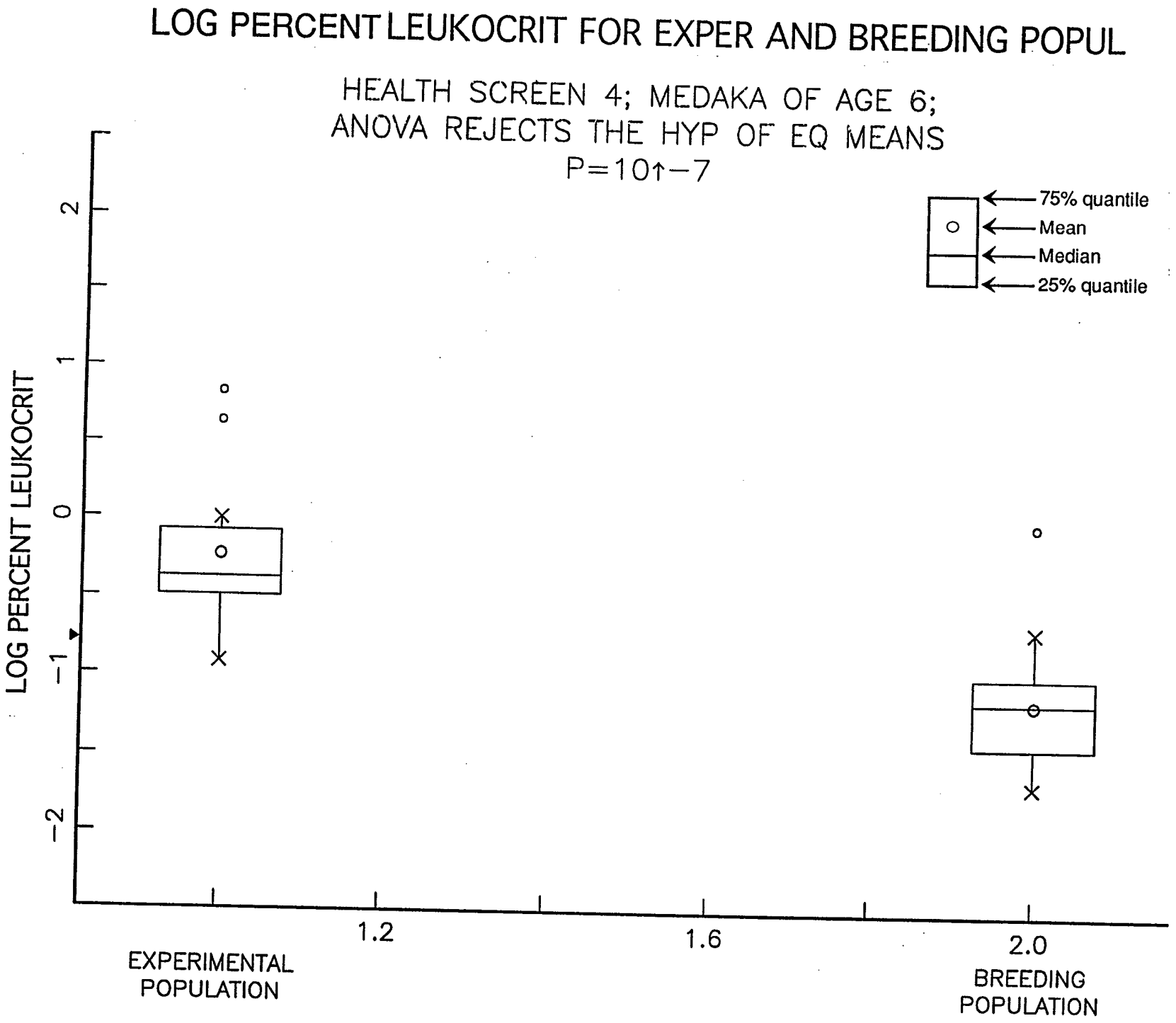

LOG PERCENT LEUKOCRIT FOR EXPER AND BREEDING POPUL
HEALTH SCREEN 4; MEDAKA OF AGE 6;
ANOVA REJECTS THE HYP OF EQ MEANS
P=10↑−7
75% quantile
Mean
Median
25% quantile
LOG PERCENT LEUKOCRIT
2
1
0
−1
−2
1.2
1.6
2.0
EXPERIMENTAL POPULATION
BREEDING POPULATION

Figure 14

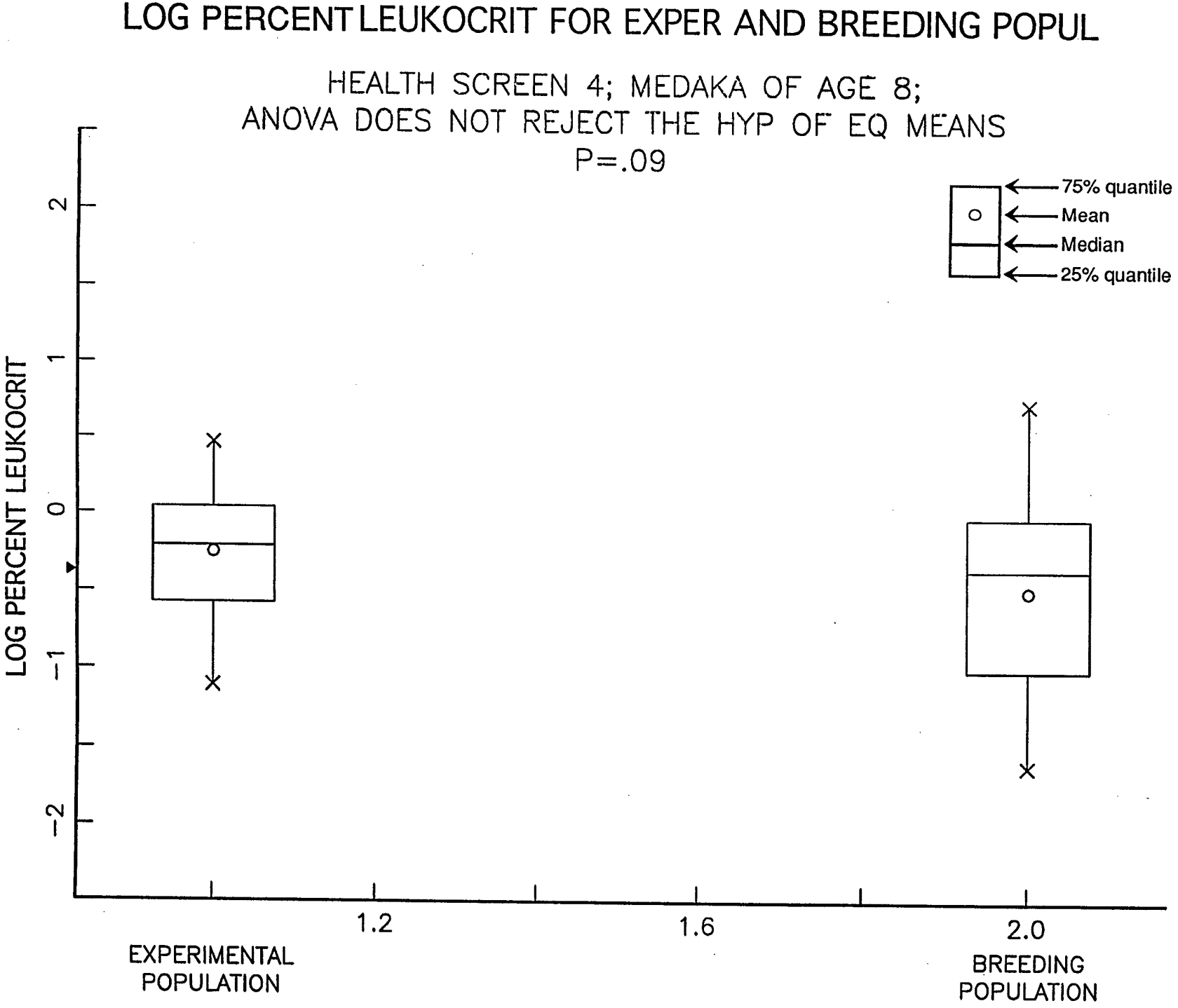

LOG PERCENT LEUKOCRIT FOR EXPER AND BREEDING POPUL
HEALTH SCREEN 4; MEDAKA OF AGE 8;
ANOVA DOES NOT REJECT THE HYP OF EQ MEANS
P=.09
75% quantile
Mean
Median
25% quantile
LOG PERCENT LEUKOCRIT
2
1
0
−1
−2
1.2
1.6
2.0
EXPERIMENTAL POPULATION
BREEDING POPULATION

Figure 15

Routine Health Monitoring in an Aquatic Species (Oryzias Latipes) Used in Toxicological Testing II:

Nonmeasurable Leukocrit Values in Health Screens 2, 3, and 4, Using Data Obtained 4/20/95 from L. Twerdok

by

P. A. Jacobs and D. P. Gaver

1. Introduction

The data consist of measurements made on Japanese medaka (Oryzias Latipes) that were sacrificed at different times during 3 health screens. Health screen 2 occurred during 7/94; health screen 3 occurred during 11/94; and health screen 4 occurred during 1/95.

The information recorded for each fish includes: the date of the experiment (which is called the sacrifice date here); the age (in months); the length (in millimeters); the weight (in milligrams); percent hematocrit; and percent leukocrit. The minimum reported value of leukocrit is 0.01 but this value is a code for "unable to measure". There are missing values which are coded by the value 100.

Previous analyses of the data reported in Gaver and Jacobs (1995) indicate an association between the ability to measure leukocrit and the length of the fish; it appears that it is more difficult to measure leukocrit in shorter fish. A weak

association was also found between the length of a fish and the measured leukocrit values; higher leukocrit values being associated with shorter fish. It may be that those fish for which leukocrit *could not* be measured had *smaller* leukocrit values. In this event, the fish for which leukocrit values *can* be measured will give a biased sample of the leukocrit values; their leukocrit values may be larger than usual; this may provide an explanation for the association of higher leukocrit values with shorter fish. In this case the missing data (unable to measure leukocrit) mechanism is said to be non-ignorable; cf. Little and Rubin (1987). A biased sample of larger than usual leukocrit values may suggest that the fish are stressed when they aren't; in an extreme case, unnecessary changes in the procedures used to culture and maintain medaka may be instituted. Dr. L. Twerdok requested that we investigate the possibility of association between the ability to measure leukocrit and the value of the leukocrit. Statistical models and methodology that address the non-ignorable missing data mechanism that may have resulted in the nonmeasurable leukocrit values have been little studied; cf. Little and Rubin (1987).

The fish used in the health screens come from several populations. One population consists of fish to be used in immunotox experiments; these fish will be called *experimental*. Another population consists of fish used for breeding; these fish will be called *breeding*. A third population consists of *retired breeding* stock fish.

In this report we present the results of analyses using statistical models to assess the association between being able to measure leukocrit and the value of leukocrit and other covariates. In Sections 2 and 3 exploratory techniques are used. In Section 4 a statistical model is proposed and maximum likelihood estimates of the parameters obtained.

The results suggest that the ability to measure leukocrit may be associated with the leukocrit value but the association does not appear to be a large effect. Analysis of data from additional health screens and biological insight are needed to determine a reasonable approach to analyzing data which have nonmeasurable leukocrits. One procedure may be to simply omit those leukocrit values that are not measurable if it is determined that this procedure will not lead to biased results. The ability to measure leukocrit is associated with the weight of the fish; (an analysis of variance gives p-value = 7.6×10^{-6}, $F = 23.2$, $df_B = 1$, $df_W = 25$). It is more difficult to measure leukocrit in fish that weigh less. There are associations between leukocrit value and the weight of the fish; (parameter estimates for regression coefficients are more than two standard deviations away from 0); this association could be due to tank effect; larger leukocrit values are associated with fish that weigh less.

2. Associations Between the Ability to Measure Leukocrit and Other Measurements

In this section we investigate possible associations between the ability to measure leukocrit and the log weight of the fish.

Figure 1 displays two boxplots of log weights. The left hand one is for animals whose leukocrit values could be measured and the right hand one is for animals whose leukocrit values could not be measured. The lengths of the two boxplots are about the same, suggesting that the variances of the *log* weights are not significantly different in the population of fish whose leukocrit level could be measured and the population in which the leukocrit level could not be measured. An Analysis of Variance rejects the null hypothesis of equal mean log weights. Thus the mean log weight for fish whose leukocrit level could not be measured is

significantly smaller ($p = 7.6 \times 10^{-6}$, $F = 23.2$, $df_B = 1$, $df_W = 25$) than the mean for fish whose leukocrit level could be measured.

The following probit model is estimated

$$P\{\text{leukocrit can } \underline{\text{not}} \text{ be measured}\} = \Phi(\beta_0 + \beta_1 \log \text{weight})$$

where Φ is the standard normal distribution. The parameter estimates are

Estimates of Probit Model

Parameter	β_0	β_1
Estimates	6.3	-1.3
(std. error)	(1.7)	(0.3)

The estimated model parameters suggest that it is harder to measure leukocrit in animals that weigh less. The slope estimate of β_1 is significantly negative (since the interval $\left[\hat{\beta}_1 - (2 \text{ std. error}) = -1.9, \hat{\beta}_1 + (2 \text{ std. error}) = -0.7\right]$ does not enclose 0).

Figure 2 presents a scatter plot of log leukocrit versus log weight for those animals whose leukocrit value *could* be measured. The line in the figure is the least squares line whose equation appears on the figure. The numbers in parentheses are the standard errors of the estimates. Note that lower leukocrit levels are associated with larger weights and the slope of the estimated straight line is significantly negative. This association could be the result of biased sampling; those smaller fish whose leukocrit values can be measured may have higher than usual leukocrit values. This conjecture will be investigated in the next sections.

Table 1 displays the estimates of fitting the probit model

$$P\{\text{measure leukocrit}\} = \Phi(\beta_0 + \beta_1(\log \text{weight}))$$

for each age of fish for which there are unmeasurable values. All of the estimates of β_1 are positive. However, all of the 95% normal confidence intervals for β_1

would include 0 suggesting no strong association between the ability to measure leukocrit and weight of fish given age.

Table 1

$P\{\text{measure leukocrit}\} = \Phi(\beta_0 + \beta_1(\log \text{weight}))$

Age	Estimates (Std. Error) β_0	β_1
3	-6.1 (7.3)	1.20 (1.4)
4	-12.5 (17.8)	2.52 (3.2)
5	-8.5 (5.1)	1.60 (0.9)
6	-8.9 (5.3)	1.75 (0.9)
8	-0.85 (3.7)	0.31 (0.6)
9 ↓	all measurable	

Table 2 displays estimates of fitting the probit model using all of the data

$$P\{\text{measure leukocrit}\} = \Phi\left(\beta_0 + \sum_i \beta_i x_i\right)$$

for various covariates x_i. All of the 3 models have about the same mean residual; this suggests that all the models summarize the data equally well. Note that the three models have estimates of β_i which are greater than 2 standard errors away from 0. Thus all of the models suggest an association between the covariates and the ability to measure leukocrit. The association may be due to the fact that the data used to estimate these models include the older and bigger fish for which all leukocrit values could be measured.

Table 2
All Data

$P\{\text{measure leukocrit}\} = \Phi(\beta_0 + \beta_1 \log \text{wt} + \beta_2 \text{ age})$				
	β_0	β_1	β_2	$\text{Mean}[\lvert O_i - \text{Fit}\rvert]$
Est.	-5.0	0.95	0.08	0.26
Std. Error	(1.9)	(0.36)	(0.05)	
$P\{\text{measure leukocrit}\} = \Phi(\beta_0 + \beta_2 (\text{age}))$				
	β_0		β_2	$\text{Mean}[\lvert O_i - \text{Fit}\rvert]$
Est.	-0.20		0.17	0.27
Std. Error	(0.34)		(0.05)	
$P\{\text{measure leukocrit}\} = \Phi(\beta_0 + \beta_1 \log \text{wt})$				
	β_0	β_1		$\text{Mean}[\lvert O_i - \text{Fit}\rvert]$
Est.	-6.34	1.27		0.26
Std. Error	(1.67)	(0.29)		

$$O_i = \begin{cases} 1 & \text{if leukocrit is measurable in fish } i \\ 0 & \text{otherwise} \end{cases}$$

3. Associations Between Log Weight and Log Leukocrit

In this section we report results of an exploratory analysis to explore possible associations between log weight and log leukocrit.

Table 3 reports the results of least squares estimation of the linear relation

$$\log \text{leukocrit} = a + b(\log \text{weight})$$

by age and population; fish without measured leukocrit are omitted. The standard errors of the estimates appear below in parentheses. Those slope estimates that are significantly different from 0 have an * beside them. For the experimental population only the regression for the fish of age 8 months has a significant slope (the 95% normal confidence interval does not include 0); the slope is significantly negative. Figure 3 (respectively Figure 4) displays scatter

plots by health screen of log weight versus log leukocrit for experimental fish of age 6 (respectively age 8). The scatter plot in the upper left hand corner displays the values for health screen 2; the scatter plot in the upper right hand corner displays the values for health screen 3; and the lower scatter plot displays the values for health screen 4. Note that for age 8 fish, there appears to be a difference in the association between log leukocrit and log weight between health screens; this could be due to tank effects.

Table 3
Least Square Straight Line Fits
log leukocrit = a + b(log weight)

	Experimental Population				Breeding Population			
Age	# fish	Intercept a (SE)	Slope b (SE)	95% Confidence Interval for b	# fish	Intercept a (SE)	Slope b (SE)	95% Confidence Interval for b
3	5	-4.3 (4.7)	0.87 (0.92)	[-2.1, 3.8]		one fish	one fish	
4	7	-4.0 (5.1)	0.70 (0.93)	[-1.7, 3.1]	6	3.8 (8.7)	-0.64 (1.5)	[-4.9, 3.6]
5	16	-1.0 (2.8)	0.20 (0.51)	[-.9, 1.3]	6	6.4 (2.3)	-1.1 (0.4)	[-2.2, 0.015]
6	25	5.2 (2.6)	-0.92 (.45)	[-1.8, 0.002]	30	7.5 (2.3)	*-1.4 (0.4)	[-2.2, -0.63]
8	36	4.6 (1.9)	*-0.83 (0.31)	[-1.5, -0.2]	44	1.6 (2.2)	-0.34 (0.38)	[-1.1, 0.43]
9	5	-3.6 (6.9)	0.66 (1.2)	[–3.3, 4.6]		no data ↓	no data ↓	
12	11	2.2 (5.0)	-0.41 (0.82)	[–2.3, 1.5]		↓ ↓	↓ ↓	
19	13	1.1 (4.1)	-0.21 (0.67)	[–1.7, 1.3]		↓ ↓	↓ ↓	

* = significant slope: the 95% confidence interval does not include 0.

The results displayed in Table 3 indicate that for the breeding population, only the regression for fish of age 6 months has a significantly negative slope. Figure 5 (respectively Figure 6) displays scatter plots of log weight versus log leukocrit by health screen for age 5 months (respectively, 6 months) fish. The left hand scatter plot of Figure 5 is for health screen 2; the right hand scatter plot is for health screen 4. The left hand scatter plot of Figure 6 is for health screen 3 and the right hand scatter plot is for health screen 4. Note that Figure 6 suggests that there is a difference in the association between log weight and log leukocrit for the two health screens; the difference could be a tank effect. Figure 5 suggests that the negative slope found in age 5 breeding population is due to 2 data points (out of 4) in health screen 3.

Tables 4 – 6 display the estimated correlations between measured log leukocrit and log weight by age of fish; those fish whose leukocrit values could not be measured are omitted. The only significant correlations for the

Table 4
Correlation Between Log Leukocrit and Log Weight
For Measured Values of Log Leukocrit By Age
All Populations

Age	Correlation	Number of Data Points	95% Confidence Interval for Correlation	
			Low	High
3	0.13	6	-0.76	0.85
4	0.09	13	-0.49	0.61
5	-0.13	22	-0.52	0.31
*6	-0.54	55	-0.70	-0.31
*8	-0.22	80	-0.42	-0.010
9	0.29	5	-0.80	0.93
12	-0.16	11	-0.69	0.48
19	-0.10	13	-0.61	0.48
* = Confidence Interval does not include 0.				

Table 5
Correlation Between Log Leukocrit and Log Weight By Age
(omits fish with no leukocrit measurement)
Experimental Population

Age	Correlation	Number of Data Points	95% Confidence Interval for Correlation	
			Low	High
3	0.48	5	-0.70	0.96
4	0.32	7	-0.57	0.86
5	0.10	16	-0.41	0.57
*6	-0.40	25	-0.68	-0.0001
*8	-0.41	36	-0.65	-0.10
9	0.29	5	-0.80	0.93
12	-0.16	11	-0.69	0.48
19	-0.10	13	-0.61	0.48
* = Confidence Interval does not include 0.				

Table 6
Correlation Between Log Leukocrit and Log Weight By Age
(omits fish with no leukocrit measurement)
Breeding Population

Age	Correlation	Number of Data Points	95% Confidence Interval for Correlation	
			Low	High
3	—	1	—	—
4	-0.20	6	-0.87	0.73
5	-0.81	6	-0.98	0.01
*6	-0.57	30	-0.77	-0.26
8	-0.14	44	-0.42	0.17
9	—	0	—	—
12	—	0	—	—
19	—	0	—	—
* = Confidence Interval does not include 0.				

experimental population are for those fish of age 6 and those fish of age 8. Note that the correlation for fish of age 6 is barely significantly negative. The correlation for fish of age 8 is significantly negative; the 95% confidence interval does not cover 0. The only significantly non-zero correlation for the breeding population is for fish of age 6 months; the correlation is significantly negative.

Table 7 records the means and variances of log weight for fish of age 6 and 8 months by population. Table 7 also records the median and interquartile range (75% quantile – 25% quantile) of the log leukocrit values. The median and interquartile range are chosen for log leukocrit since they are robust to nonmeasurable values. There are two sets of statistics for each age. Those statistics in the columns labeled *all* assume that the nonmeasurable leukocrit values are all smaller than those that are measurable. Those statistics in the columns labeled *measurable* use only the measurable leukocrit values.

Table 7
Descriptive Measures of Location and Spread
All Fish by Age

Age	Popul.	# Missing Leukocrit	# Fish	Log Weight		Log Leuk All (Missing Values Assumed Small)		Log Leuk Measurable	
				Mean	Var	Median	$Q_{.75}$–$Q_{.25}$ (Est of log std dev)	Median	$Q_{.75}$–$Q_{.25}$ (Est of log std dev)
6	Exp.	6	31	5.72	0.031	-0.19	0.82 (-0.49)	-0.11	0.67 (-0.69)
6	Breed.	2	32	5.82	0.078	-1.12	0.72 (-0.62)	-1.07	0.67 (-0.69)
8	Exp.	7	43	5.94	0.056	-0.26	0.85 (-0.46)	-0.22	0.63 (-0.76)
8	Breed.	8	52	5.89	0.067	-0.43	1.35 (0.001)	-0.35	0.72 (-0.62)

To obtain a robust estimate of spread note that the interquartile distance for a standard normal is 1.348. Let Z be a standard normal random variable; then

$$0.25 = P\{\sigma Z \le q_{.25}\} = P\left\{Z \le \frac{1}{\sigma} q_{.25}\right\}$$

where $q_{.25}$ is the 0.25 quantile of a normal random variable with mean 0 and variance σ^2. Thus,

$$\frac{1}{\sigma}[q_{.75} - q_{.25}] = 1.348$$

or

$$\hat{\sigma} = \frac{q_{0.75} - q_{0.25}}{1.348}$$

is an estimate of the standard deviation of a normal random variable. The values in parentheses below the interquartile distances in Table 7 are the estimates of the log standard deviation $\log \hat{\sigma} = \log\left[\frac{\hat{q}_{0.75} - \hat{q}_{0.25}}{1.348}\right]$. The median of log leukocrit is an estimate of its mean if it is assumed that log leukocrit is normally distributed.

Note that those estimates for the median and log standard deviation for log leukocrit are always more extreme if one assumes all the nonmeasurable leukocrit values are smaller than those that could be measured. In the next section we use statistical models to assess the effect of nonmeasurable leukocrit values on the summary statistics of log leukocrit.

4. Results of a Model to Assess the Effect of Nonmeasurable Leukocrit Values on Estimates of Moments Involving Log Leukocrit

In this section we introduce a model to assess the effect of the nonmeasurable leukocrit values on the moment estimates involving log leukocrit. One possible effect is as follows. If a leukocrit value is not measurable because it is smaller than those that could be measured, then the mean log leukocrit value obtained by averaging those that could be measured will be too high which may suggest that

the fish are stressed when in fact they aren't. Further, association between leukocrit and other variables may also be distorted. We will call those data for which leukocrit values cannot be measured censored.

The model is described in detail in Appendix A. It consists of two parts. The pairs $\{(Y_i, W_i)\}$ of log leukocrit and log weight for a fixed age of fish are assumed to have a bivariate normal distribution. Further,

$$P\{\text{leukocrit for fish } i \text{ is measurable} \mid Y_i = y, W_i = w\} = \Phi(ay + bw + c)$$

where $\Phi(x)$ is the cumulative distribution of a standard normal distribution; that is, the probability of being able to measure leukocrit is described by a probit model with covariates log leukocrit and log weight. We obtain estimates for two models. In one $a = 0$ is fixed; that is, the ability to measure leukocrit is only a function of the log weight of the animal. The other model also estimates a; that is, the model allows the possibility that the ability to measure leukocrit is also a function of the value of the leukocrit.

Tables 8 – 9 display maximum likelihood estimates and standard errors of the moments of the bivariate normal distribution model for log leukocrit and log weight for experimental and breeding populations under the two probit censoring models. Table 8 displays results for age 6 month and 8 month medaka for the probit censoring model in which the probability of being able to measure leukocrit is a function only of the value of log weight. Table 9 displays results for the censoring model in which the probability of being able to measure leukocrit is a function of both the value of the log leukocrit and log weight. Note the extreme values for the probit estimates in Table 9 for the probit censoring model that includes leukocrit for the age 6 month medaka. Also note the large standard errors in Table 9 associated with the probit parameter estimates for the censoring

Table 8

Estimates for the Moments of Joint Distribution of Log Weight and Log Leukocrit

The probability of being able to measure leukocrit is a function of log weight.

			Estimated Parameters (Standard Errors)						
			Probit		Bivariate Normal				
					Mean		log std. dev.		Corr.
Age	Population	# of Fish	log wt b	constant c	log leuk. m_1	log wt m_2	log leuk. τ_1	log wt τ_2	ρ
6	Experimental	31	2.38 (1.83)	-12.7 (10.4)	-0.08 (0.08)	5.72 (0.02)	-0.872 (0.14)	-1.75 (0.13)	-0.381 (0.15)
6	Breeding	32	1.51 (1.27)	-7.12 (7.19)	-0.798 (0.126)	5.82 (0.047)	-0.351 (0.124)	-1.29 (0.12)	-0.563 (0.10)
8	Experimental	43	0.076 (0.97)	0.53 (5.77)	-0.30 (0.074)	5.94 (0.025)	-0.766 (0.117)	-1.45 (0.107)	-0.417 (0.119)
8	Breeding	52	0.507 (0.826)	-1.96 (4.85)	-0.389 (0.094)	5.89 (0.025)	-0.465 (0.107)	-1.36 (0.10)	-0.138 (0.15)

model that includes the value of log leukocrit. These large standard errors suggest that the likelihood is very flat around the estimates. This could be due to the small sample sizes and large number of parameters to be fit. It could also indicate that the data do not provide clear indication of the association between log leukocrit and the ability to measure log leukocrit. This lack of clear indication is also suggested by the estimates of the mean log leukocrit in Table 9; note that they are smaller than those for the age 6 month medaka in Table 8 but larger than those for the age 8 month medaka in Table 8. Thus, the model suggests that the leukocrit values that could not be measured for age 6 month medaka tend to be smaller than those that could. However, the model suggests that leukocrit values that could not be measured for the age 8 medaka are not necessarily smaller than those that could be measured.

Table 9

Estimates for the Moments of the Joint Distribution of Log Weight and Log Leukocrit

The probability of being able to measure leukocrit is a function of log weight and log leukocrit.

				Estimated Parameters (standard errors)							
				Probit			Bivariate Normal				
							Mean		log std. dev.		Correl.
Age	Population	# of Fish	# without leukocrit measurement	log leuk. a	log wt b	constant	log leuk. m_1	log wt m_2	log leuk. τ_1	log wt. τ_2	ρ
6	Experimental	31	6	1×10^4 (1×10^5)	3×10^4 (3×10^5)	-2×10^5 (3×10^5)	-0.25 (0.09)	5.7 (0.02)	-0.67 (0.14)	-1.75 (0.13)	-0.17 (0.18)
6	Breeding	32	2	1×10^3 (8×10^3)	2×10^3 (1×10^4)	-1×10^4 (7×10^5)	-0.87 (0.13)	5.8 (0.05)	-0.34 (0.13)	-1.29 (0.12)	-0.48 (0.12)
8	Experimental	43	7	-6.0 (3.6)	-2.6 (3.3)	17.8 (19.6)	-0.16 (0.09)	5.9 (0.03)	-0.61 (0.13)	-1.45 (0.11)	-0.34 (0.13)
8	Breeding	52	8	-1.6 (1.6)	0.02 (1.1)	1.0 (6.9)	-0.27 (0.14)	5.9 (0.02)	-0.35 (0.18)	-1.4 (0.1)	-0.16 (0.14)

Table 10 displays three estimates of the moments of the bivariate normal distribution model for log leukocrit and log weight. The table reports the sample moments; the maximum likelihood estimates (MLE) resulting from the probit censoring model in which the probability of being able to measure leukocrit depends only on the value of log weight, and the maximum likelihood estimates resulting from the censoring model in which the probability of being able to measure leukocrit depends on the value of the log leukocrit and the log weight. The sample moments involving log leukocrit are computed by leaving out the missing values. Note that the sample moment estimates and the maximum likelihood estimates using the probit censoring model with only log weight are about the same. The maximum likelihood estimates using the probit model which includes log leukocrit are the same as the others for the moments of log weight and are within two standard errors of the others for the moments involving log leukocrit. Comparing the maximum likelihood estimates of the mean log leukocrit to the two median log leukocrit estimates appearing in Table 7, note that both median estimates fall within 2 standard errors of the mean estimates except for the age 6 month breeding population. In this case the median which is computed by assuming all the nonmeasurable leukocrit values are smaller than those that could be measured falls outside 2 standard errors of the MLE estimate using the probit censoring model with only log weight; however, it is within 2 standard errors of the MLE using the probit censoring model with log leukocrit and log weight.

Note that the correlations for the age 6 breeding population and the age 8 experimental population are more than 2 standard errors away from 0 for both models and the sample moments; (see Tables 5 and 6). This suggests that heavier

Table 10

Estimates of the Moments of the Joint Distribution of Log Leukocrit and Log Weight

			Estimates (std. error)				
			Mean		log std. dev.		Correl.
Age	Population	Method/Model	Log Leuk.	log wt.	Log Leuk.	log wt	
6	Experimental	Moment	-0.10 (–)	5.72 (–)	-0.85 (–)	-1.74 (–)	-0.40 (–)
		MLE, measurability = f(log wt)	-0.08 (0.08)	5.72 (0.02)	-0.87 (0.14)	-1.75 (0.13)	-0.38 (0.15)
		MLE, measurability = f(log wt, log leuk.)	-0.25 (0.09)	5.72 (0.02)	-0.67 (0.15)	-1.75 (0.13)	-0.17 (0.18)
6	Breeding	Moment	-0.82 (–)	5.82 (–)	-0.33 (–)	1.28 (–)	-0.57 (–)
		MLE, measurability = f(log wt)	-0.80 (0.13)	5.82 (0.05)	-0.35 (0.12)	-1.29 (0.12)	-0.56 (0.10)
		MLE, measurability = f(log wt, log leuk.)	-0.87 (0.13)	5.82 (0.05)	-0.34 (0.13)	-1.29 (0.12)	-0.48 (0.12)
8	Experimental	Moment	-0.30 (–)	5.94 (–)	-0.75 (–)	1.44 (–)	-0.41 (–)
		MLE, measurability = f(log wt)	-0.30 (0.07)	5.94 (0.03)	-0.77 (0.12)	-1.45 (0.11)	-0.42 (0.12)
		MLE, measurability = f(log wt, log leuk.)	-0.16 (0.09)	5.94 (0.03)	-0.61 (0.13)	-1.45 (0.11)	-0.34 (0.13)
8	Breeding	Moment	-0.39 (–)	5.89 (–)	-0.45 (–)	1.35 (–)	-0.14 (–)
		MLE, measurability = f(log wt)	-0.40 (0.09)	5.89 (0.03)	-0.47 (0.11)	-1.36 (0.10)	-0.14 (0.15)
		MLE, measurability = f(log wt, log leuk.)	-0.25 (0.14)	5.89 (0.02)	-0.35 (0.18)	-1.36 (0.10)	-0.16 (0.14)

fish have lower leukocrit values for these populations; this association could be due to tank effect.

Note also that the correlations estimated using moments (see Table 5) and the probit model with only log weight are negative and more than two standard deviations away from 0 for the experimental population of age 6 medaka. However, the correlation estimated for the experimental population of age 6 medaka using the probit censoring model with log leukocrit and log weight is negative and within two standard deviations of the origin. Further, the estimate of mean log leukocrit is lower for the probit model with log leukocrit and log weight than the other two estimation procedures. Thus, the more elaborate censoring model is suggesting that the nonmeasurable leukocrit values are smaller than the others for this case. However, note that the mean log leukocrit values estimated using the probit model with log leukocrit and log weight are larger than those for the other two procedures for the age 8 medaka.

Appendix B displays results for another model to assess the effect of the nonmeasurable leukocrit values.

5. Conclusions

The ability to measure leukocrit is associated with log weight; the larger the log weight, the greater the probability of being able to measure leukocrit. The value of measured log leukocrit is associated with weight of the fish for ages 6 and 8 month medaka. This association could be due to tank effects. Any association between the ability to measure leukocrit and the leukocrit value itself appears to be small.

REFERENCES

D. P. Gaver and P. A. Jacobs, "Routine Health Monitoring in an Aquatic Species (Oryzias Latipes) Used in Toxicological Testing I: Preliminary Examination of Leukocrit Data from Health Screens 2, 3, and 4, Using Data Obtained 4/20/95 from L. Twerdok," Working Paper, April 1995.

IBM Corporation. A Graphical Statistical System (AGSS).

R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley, New York, 1987.

APPENDIX A
A Bivariate Normal Model with Censoring

In this Appendix we present details of a bivariate normal model with missing values.

Let $\{(Y_i, W_i)\}$ be independent bivariate normal random variables with $E[Y_i] = m_1$, $E[W_i] = m_2$, $\text{Corr}(Y_i, W_i) = \rho$, $Var[Y_i] = e^{2\tau_1}$ and $Var[W_i] = e^{2\tau_2}$. For the health screen data, Y_i = log leukocrit, W_i = log weight.

Assume for each fish i there is a random tolerance Z_i. Assume the leukocrit for fish i is <u>not</u> measurable if a linear combination of log leukocrit and log weight falls below the threshold; that is, assume Y_i is not observable if $aY_i + bW_i + c < Z_i$. Assume the random tolerances $\{Z_i\}$ are iid standard normal.

Let $R_i = 1$ if the log leukocrit Y_i is observable and $R_i = 0$ if it is not.

$$P\{R_i = 0 \mid Y_i = y, W_i = w\} = P\{ay + bw + c < Z_i\} = 1 - \Phi(ay + bw + c)$$

where Φ is the standard normal cumulative distribution function

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz$$

and $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}$ is the density function of a standard normal random variable.

The conditional distribution of log leukocrit Y_i given log weight $W_i = w$ is normal with mean $m_1 + \rho e^{\tau_1 - \tau_2}(w - m_2)$ and variance $e^{2\tau_1}\left(1 - \rho^2\right)$. Thus, letting $\tilde{Z}$ be another independent standard normal random variable

$$P\{R_i = 0 | W_i = w\} = P\{aY_i + bw + c < Z_i | W_i = w\}$$

$$= P\left\{a\left[m_1 + \rho e^{\tau_1 - \tau_2}(w - m_2) + e^{\tau_1}\sqrt{1-\rho^2}\tilde{Z}\right] + bw + c < Z_i | W_i = w\right\}$$

$$= 1 - \Phi\left(\left\{a\left[m_1 + \rho e^{\tau_1 - \tau_2}(w - m_2)\right] + bw + c\right\}\left[1 + a^2 e^{2\tau_1}\left(1-\rho^2\right)\right]^{-1/2}\right).$$

Let

$$h(\boldsymbol{\theta}; w) \equiv h(a,b,c,m_1,m_2,\tau_1,\tau_2,\rho;w) = \frac{a\left[m_1 + \rho e^{\tau_1 - \tau_2}(w - m_2)\right] + bw + c}{\sqrt{1 + a^2 e^{2\tau_1}\left(1-\rho^2\right)}}.$$

The log likelihood function for the model is up to addition of constants

$$\ell = \sum_i (1 - r_i)\left\{\log\left[1 - \Phi\left(h(\boldsymbol{\theta}; w_i)\right)\right] - \tau_2 - \frac{1}{2}(w_i - m_2)^2 e^{-2\tau_2}\right\}$$

$$+ \sum_i r_i \left\{\log\Phi(ay_i + bw_i + c) - \tau_1 - \tau_2 - \frac{1}{2}\log\left(1-\rho^2\right)\right.$$

$$\left. - \frac{1}{2\left(1-\rho^2\right)}\left[\frac{(y_i - m_1)^2}{e^{2\tau_1}} - 2\rho\frac{(y_i - m_1)}{e^{\tau_1}}\frac{(w_i - m_2)}{e^{\tau_2}} + \frac{(w_i - m_2)^2}{e^{2\tau_2}}\right]\right\}.$$

Note that when $a = 0$, the log-likelihood simplifies to a sum of a separate probit log likelihood and the log-likelihood for the moments of the bivariate normal. As a result the parameters are estimated as follows.

1. Fix $a = a_0$.
2. Estimate the other parameters, b, c, m_1, m_2, τ_1, τ_2, ρ, applying Newton-Raphson using ℓ with fixed $a = a_0$.
3. Evaluate the full likelihood function and its first derivatives at a_0 and the estimates found in 2.
4. Choose a new value for a and go back to 2.

The maximum likelihood estimate is found by first searching on a grid for a and then applying the Newton-Raphson procedure on all the parameters when the partial derivative of the log-likelihood with respect to a is close enough to 0.

APPENDIX B

In this Appendix we present another model for assessing the sampling effect of censoring. This model has found application in the econometrics literature.

B.1 A Bivariate Normal Stochastic Censoring Model

Let Y_{2i} be a measure of the difficulty of measuring leukocrit in fish i and let Y_{1i} be the log leukocrit level in fish i. Assume $\{(Y_{1i}, Y_{2i})\}$ are independent bivariate normal random variables with $\text{Var}[Y_{1i}] = \sigma_{11}$, $\text{Var}[Y_{2i}] = \sigma_{22}$ and $\text{Cov}[Y_{1i}, Y_{2i}] = \sigma_{12}$; $E[Y_{i1}] = x_i\beta(1) \equiv \sum_{j=1}^{p_1} x_{ij}(1)\beta_j(1)$; $E[Y_{i2}] = x_i\beta(2) \equiv \sum_{j=1}^{p_2} x_{ij}(2)\beta_j(2)$ where $x_i(1) = \left(x_{i1}(1),\ldots,x_{ip_1}(1)\right)$ and $x_i(2) = \left(x_{i1}(2),\ldots,x_{ip_2}(2)\right)$ are covariates. We assume that the measure of ability to measure leukocrit Y_{2i} is never observed and Y_{1i} is observed only if $Y_{2i} > 0$. Let

$$R_i = \begin{cases} 1 & \text{if } Y_{i2} > 0 \text{ and } Y_{i1} \text{ is observed} \\ 0 & \text{if } Y_{i2} \le 0. \end{cases}$$

The likelihood function for this model is

$$L = \prod_{i=1}^{n} \Phi\left(\frac{-x_i(2)\beta(2)}{\sqrt{\sigma_{22}}}\right)^{1-r_i} \left\{\left[1-\Phi\left(\frac{-x_i(2)\beta(2)}{\sqrt{\sigma_{22}}}\right)\right] P\left\{Y_{1i} \in dy_{1i} \,\middle|\, Y_{2i} > 0\right\}\right\}^{r_i}.$$

Thus, the log likelihood function is

$$
\begin{aligned}
\ell &= \log L \\
&= \sum_{i=1}^{n}(1-r_i)\log\left[1-\Phi\left(\frac{x_i(2)\beta(2)}{\sqrt{\sigma_{22}}}\right)\right] \\
&+r_i\left[\log\Phi\left(\frac{x_i(2)\beta(2)}{\sqrt{\sigma_{22}}}\right)+\log P\left\{Y_{1i}\in dy_{1i}\,\middle|\,Y_{2i}>0\right\}\right] \\
&= \sum_{i=1}^{n}(1-r_i)\log\left[1-\Phi\left(\frac{x_i(2)\beta(2)}{\sqrt{\sigma_{22}}}\right)\right] \\
&+\sum_{i=1}^{n}r_i\left[\log\Phi\left(x_i(2)\beta(2)\frac{1}{\sqrt{\sigma_{22}}}+\frac{\sigma_{12}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}}\frac{(y_{1i}-x_i(1)\beta(1))}{\sqrt{\sigma_{11}}}\right)\right. \\
&\left.-\frac{1}{2}\log\left[1-\frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}\right]-\frac{1}{2}\log\sigma_{11}+\log\varphi\left(\frac{1}{\sqrt{\sigma_{11}}}(y_{1i}-x_i(1)\beta(1))\right)\right]
\end{aligned}
$$

where φ is the standard normal density function, cf. Amemiya (1985). This model was introduced by Heckman (1976) to describe selection of women into the labor force. Amemiya (1985) calls the model a Type II Tobit model. Heckman, in Heckman (1979), describes a simple but inefficient procedure to estimate the parameters. Little and Rubin (1987) make cautionary remarks concerning use of the procedure.

The Heckman two-step estimator is as follows. Assume the data are ordered so that the observed values of y_{i1} are the first n_1 values.

1. Estimate $\alpha=\beta(2)/\sqrt{\sigma_{22}}$ by the probit maximum likelihood estimator.
2. Regress y_{i1} on $x_i(1)$ and $\lambda(x_i(2)\hat{\alpha})$ by least squares using only the observed y_{i1} where

$$\hat{\lambda}_i \equiv \lambda(x_i(2)\hat{\alpha}) = \frac{\varphi(x_i(2)\hat{\alpha})}{\Phi(x_i(2)\hat{\alpha})}.$$

The resulting estimate is $\gamma=\left(\beta(1),\hat{C}\right)$ where $\hat{C}$ is an estimate of $\frac{\sigma_{12}}{\sqrt{\sigma_{22}}}$.

C is a measure of selection bias; that is association between the value of log leukocrit and its ability to be measured. A value of $C = 0$ indicates that there is no association between the ability to measure leukocrit and its value. To test for no selection bias; that is, the null hypothesis is $C = \sigma_{12}/\sqrt{\sigma_{22}} = 0$, a t-test can be performed using the usual regression standard error for $\hat{C}$; cf. Heckman (1979).

B.2 Application of the Bivariate Normal Stochastic Censoring Model to Health Screen Data

The parameters of the model of Section B.1 are estimated with dependent variable log leukocrit. The covariates for the probit regression are a constant and log weight; that is

$$P\{\text{being able to measure leukocrit of fish} \mid \text{log weight of fish}\}$$
$$= P\{Y_{2i} > 0\} = \Phi(\beta_{20} + (\beta_{21}(\text{log weight}_i)).$$

The covariates for the observed log leukocrit are a constant and length; that is, $x_{1i} = (1, \text{length})$.

Heckman's two-step estimator is used. The estimates appear in Table B.1. Let $\alpha_{2i} = \frac{\beta_{2i}}{\sqrt{\sigma_{22}}}$, the parameters from the probit regression.

Table B.1
Estimate for Tobit Model
All data

Covariate	Probit P(measuring leukocrit)		Log leukocrit		
	Intercept α_{20}	log weight α_{21}	Intercept β_{10}	length β_{11}	λ $\hat{C}$
Estimate	-6.34	1.27	2.09	-0.09	-0.16
(Std. Error)*	2.80	0.09	(0.83)	(0.03)	(0.48)
* Standard errors are the usual regression standard errors.					

Note that $\hat{C} \equiv \sigma_{12}/\sqrt{\sigma_{22}}$ is not significantly different than 0. Thus, there is no indication that the values of the log leukocrit are influenced by the ability to measure leukocrit.

The estimates of parameters for the model using the experimental population of age less than or equal to 8 months appear in Table B.2.

Table B.2
Experimental Population
Age less than or equal to 8 months

	Probit P(measuring leukocrit)		Regression		
	Intercept α_{20}	log weight α_{21}	Intercept β_{10}	length β_{11}	λ C
Estimate	-3.9	0.83	2.30	-0.09	-0.46
(Std. Error)*	(5.4)	(0.17)	(1.2)	(0.04)	(0.81)
* Standard errors are the usual.					

Note that for the experimental population the estimate of C is not significantly different than 0, indicating that the ability to measure leukocrit is not associated with the value of the leukocrit level.

The estimates for the model for only the breeding population of age less than or equal to 8 months appear in Table B.3.

Table B.3

Breeding Population

Age less than or equal to 8 months

	Probit P(measuring leukocrit)		Regression		
	Intercept α_{20}	log weight α_{21}	Intercept β_{10}	length β_{11}	λ C
Estimate	-8.4	1.6	2.15	-0.10	0.25
(Std. Error)*	(8.2)	(0.3)	(1.4)	(0.05)	(0.73)
* Standard errors are the usual regression standard errors.					

Note that for the breeding population, the estimate of C is not significantly different than 0 indicating that the ability to measure leukocrit is not associated with the value of the leukocrit.

REFERENCES

T. Amemiya, *Advanced Econometrics*, Harvard University Press, Cambridge, Mass. 1985.

J. J. Heckman, "The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models," *Ann. Econ. Soc. Meas.*, **5** (1976) pp. 475–492.

J. J. Heckman, "Sample selection bias as a specification error," *Econometrica*, **4** (1979) pp. 153–161.

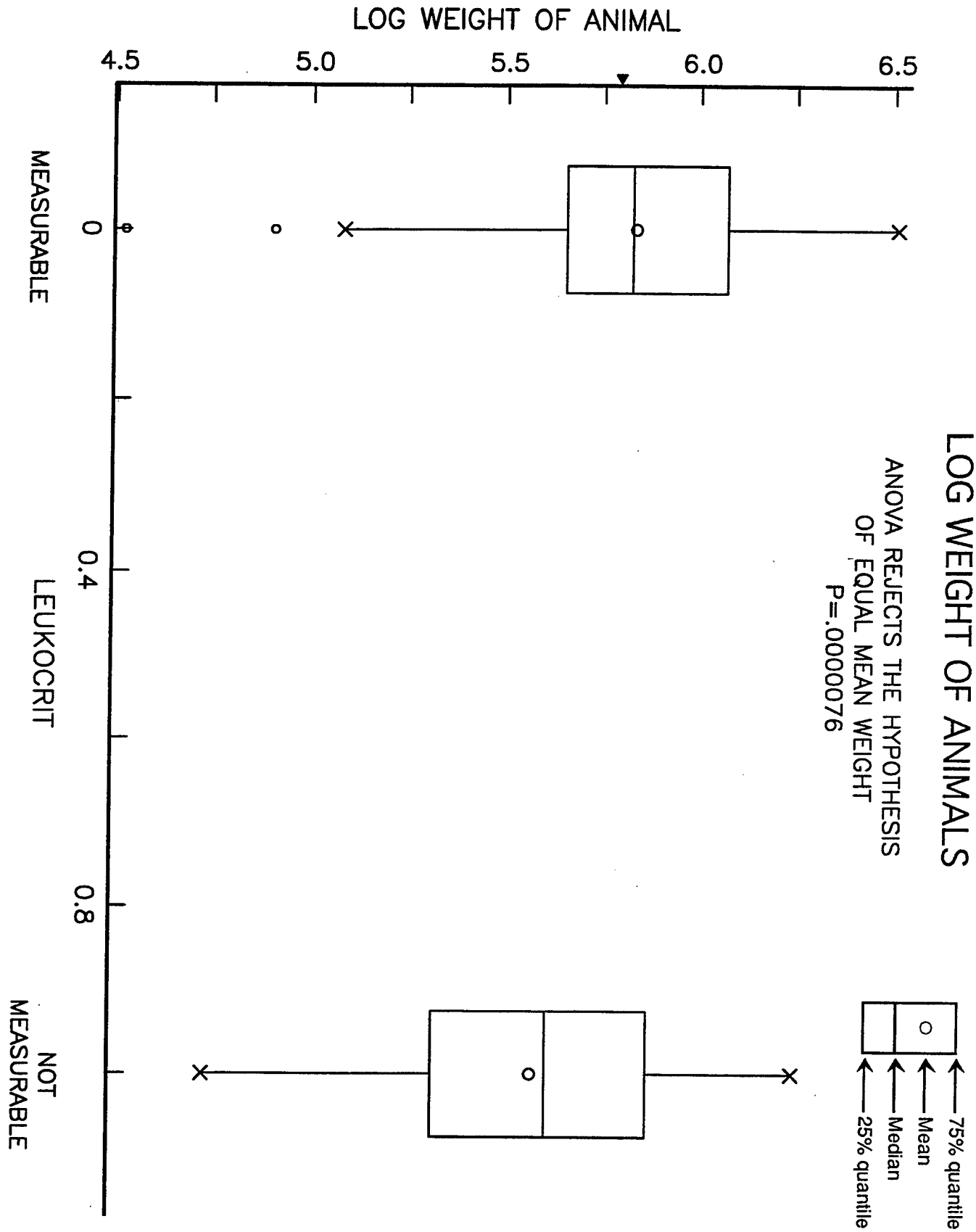

LOG WEIGHT OF ANIMAL
4.5
5.0
5.5
6.0
6.5
MEASURABLE
0
LEUKOCRIT
0.4
0.8
NOT MEASURABLE
LOG WEIGHT OF ANIMALS
ANOVA REJECTS THE HYPOTHESIS
OF EQUAL MEAN WEIGHT
P=.000076
75% quantile
Mean
Median
25% quantile

Figure 1

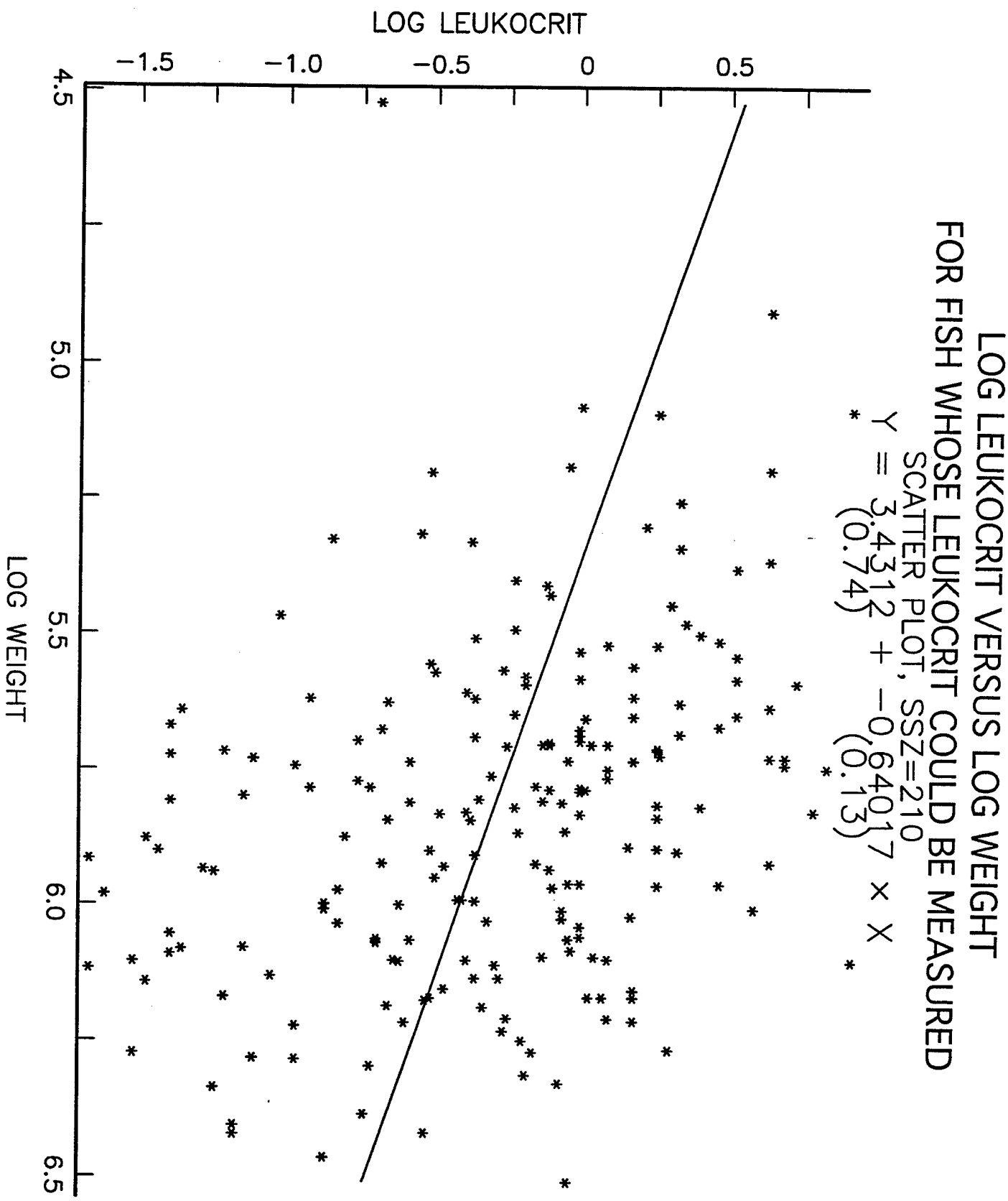

LOG LEUKOCRIT
−1.5
−1.0
−0.5
0
0.5
4.5
5.0
5.5
6.0
6.5
LOG WEIGHT
LOG LEUKOCRIT VERSUS LOG WEIGHT
FOR FISH WHOSE LEUKOCRIT COULD BE MEASURED
SCATTER PLOT, SSZ=210
Ŷ = 3.4312 + −0.6401 X
(0.74) (0.13)

Figure 2

AGE=6;EXPERIMENTAL POPULATION
((VLEU≠.01)^(VAGE=6)^(STATUS=1))/STHS=2
SSZ=4
((VLEU≠.01)^(VAGE=6)^(STATUS=1))/STHS=3
SSZ=6
((VLEU≠.01)^(VAGE=6)^(STATUS=1))/STHS=4
SSZ=15
LOG LEUKOCRIT
LOG WEIGHT
0.8
0.4
0
−0.4
−0.8
5.4
5.6
5.8
6.0
6.2

Figure 3

AGE=8; EXPERIMENTAL POPULATION
((VLEU≠.01)^(VAGE=8)^(STATUS=1))/STHS=2
SSZ=3
((VLEU≠.01)^(VAGE=8)^(STATUS=1))/STHS=3
SSZ=6
((VLEU≠.01)^(VAGE=8)^(STATUS=1))/STHS=4
SSZ=27
LOG LEUKOCRIT
LOG WEIGHT
0
−0.5
−1.0
−1.5
5.4
5.6
5.8
6.0
6.2
6.4

Figure 4

AGE=5;BREEDING POPULATION

Figure 5

AGE=6;BREEDING POPULATION

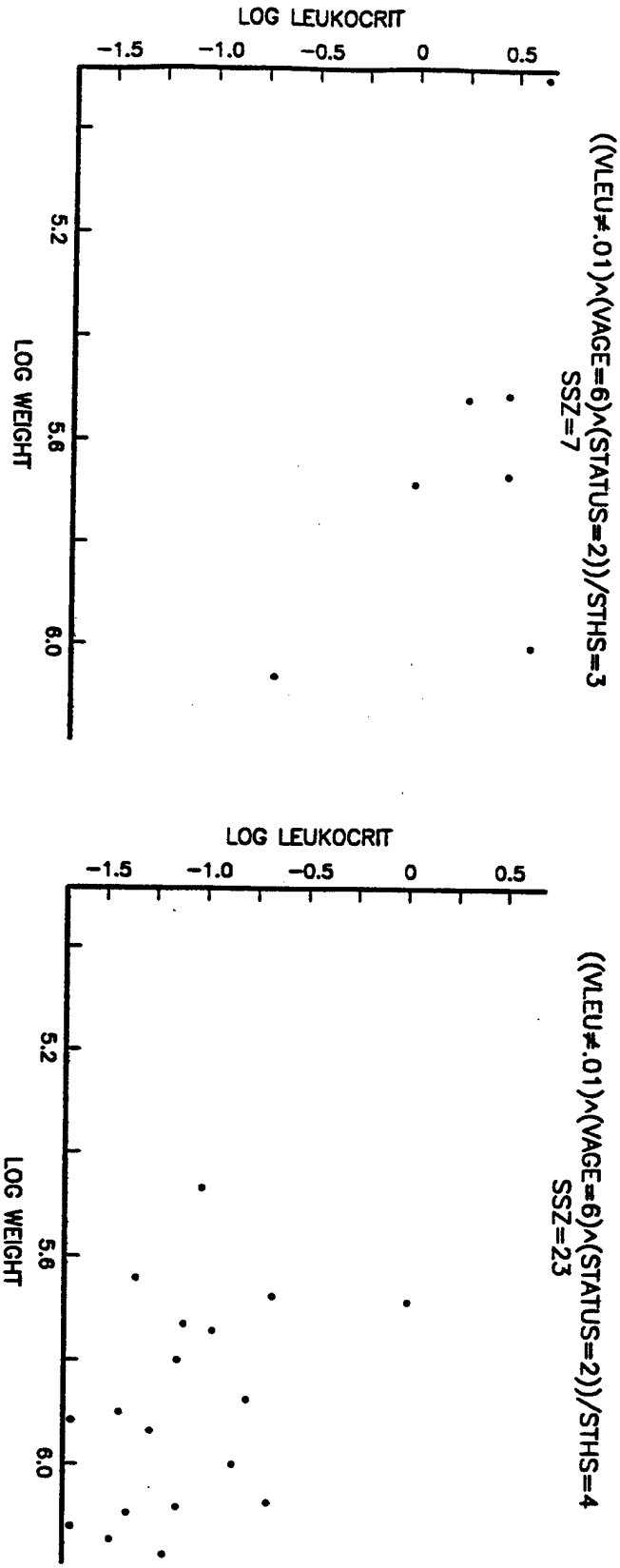


Figure 6

Routine Health Monitoring in an Aquatic Species (Oryzias Latipes) Used in Toxicological Testing III:

Exploratory Data Analysis Using Multivariate Comparison of Populations Using Data Obtained 4/20/95 from L. Twerdok

by

D. P. Gaver and P. A. Jacobs

1. Introduction

The data consist of measurements made on Japanese medaka (Oryzias Latipes) that were sacrificed at different times during 3 health screens. Health screen 2 occurred during 7/94; health screen 3 occurred during 11/94; and health screen 4 occurred during 1/95.

The information recorded for each fish includes: the date of the experiment (which is called the sacrifice date here); the age (in months); the length (in millimeters); the weight (in milligrams); percent hematocrit; and percent leukocrit. The minimum reported value of leukocrit is 0.01 but this value is a code for "unable to measure". There are missing values which are coded by the value 100.

The fish used in the health screens come from several populations. One population consists of fish to be used in immunotox experiments; these fish will

be called *experimental*. Another population consists of fish used for breeding; these fish will be called *breeding*. A third population consists of *retired breeding* fish.

Previous analyses (cf. Twerdok et al.) of the data have considered comparisons between populations using one type of measurement at a time (e.g. length). Analyses restricted to one measurement at a time may overlook differences in the association between measurements for different populations. In this paper we describe a standard statistical procedure for comparing vectors of means between two populations. This technique finds the linear combination of the measurements which results in the greatest discrepancy between the two populations; thus it implicitly considers the univariate comparisons and incorporates the variance-covariance matrix of the measurements. If a (statistically) significant difference is found, further data analysis is needed to determine the reason. Finally, the biological significance of the difference needs to be assessed.

Section 2 describes the procedure. Section 3 describes results obtained by applying the procedure to length, log weight, and log hematocrit for breeding and experimental populations of medaka that are 8 months of age. It is found that there is no statistically significant difference between the mean vectors of length, log weight and log hematocrit in the two populations. Section 3 also describes results of applying the procedure to length, log weight, and log hematocrit for breeding and experimental populations of medaka that are 6 months of age. For medaka of this age there is a significant difference in the mean vectors. Section 4 describes the results of applying the procedure to log hematocrit and log leukocrit for those fish in the breeding and experimental populations that have measured leukocrit.

2. Comparison of Multivariate Means for Two Populations

2.1 Summary Statistics

Suppose one has collected observations on p different variables $(x_{i1}, \ldots, x_{ip})$ for a number of fish $i = 1, \ldots, n$. Summary statistics for the data matrix $\mathbf{X}$ with i^{th} row $(x_{i1}, \ldots, x_{ip})$ are

1. The sample mean

   Let $\bar{x}_j = \frac{1}{n}\sum_{r=1}^{n} x_{rj}$.

   The sample mean vector, $\bar{x}$, is given by $\bar{x}^T = \left[\bar{x}_1, \ldots, \bar{x}_p\right]$.

2. The sample covariance of variables k and j is

   $$s_{kj} = \sum_{r=1}^{n} (x_{rk} - \bar{x}_k)\left(x_{rj} - \bar{x}_j\right)/(n-1)$$

   The sample covariance matrix is

   $$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

2.2 Comparison of Mean Vectors for Two Populations

Suppose one has collected observations from 2 populations and wishes to compare the vector of means from the two populations; e.g. the measurement of length, log weight, and log hematocrit from a health screen of medaka of a particular age (e.g. 8 months) for the experimental population and the breeding population. If the sample sizes are of size n_1 and n_2 respectively, then for $i = 1, 2$ the data matrix $\mathbf{X}(i)$ is of order $(n_i \times p)$ and represents a random sample of

independent observations from an *assumed* multivariate normal distribution with vector mean μ_i and variance-covariance Σ. Note that we are assuming that the two population variance-covariance matrices are the same.

A generalization of the univariate two-sample procedure is as follows; (cf. Chatfield and Collins, 1980).

1. The pooled within-groups estimate of S is given by

$$\mathbf{S} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

where $\mathbf{S_i}$ is the sample variance-covariance matrix for population *i*.

2. Compute $\mathbf{a}^* = \mathbf{S}^{-1}(\bar{\mathbf{x}}(1) - \bar{\mathbf{x}}(2))$

where $\bar{\mathbf{x}}(\mathbf{i})$ is the sample column vector mean for population *i*.

3. Compute $T^2 = \frac{n_1 n_2}{(n_1 + n_2)}(\bar{x}(1) - \bar{x}(2))^T \mathbf{a}^*$.

4. Under the null hypothesis that $\mu_1 = \mu_2$ the statistic

$$\frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

has an F distribution with numerator degrees of freedom p and denominator degrees of freedom $n_1 + n_2 - p - 1$.

The assumption that the covariance matrices of the two populations are equal is a generalization of the assumption of equal variances in the univariate case. However, the T^2-statistic is not sensitive to departures from the assumption when the sample sizes are approximately equal (cf. Chatfield and Collins [1980]). Note that since more parameters are being estimated more data are required than in the univariate case.

Suppose we consider linear combinations of the data in the two populations $U_r(i,\mathbf{a}) = \sum_{k=1}^{p} X_{rk}(i)a_k$. The vector $\mathbf{a}^*$ is that value of $\mathbf{a} = (a_1, \ldots, a_p)$ that produces the greatest inconsistency between the two populations as measured by the t-statistic used to compare the means the two population's univariate observations $\{U_r(1, \mathbf{a}), r = 1, \ldots, n_1\}$ and $\{U_r(2, a), r = 1, \ldots, n_2\}$.

3. Comparisons of length, log weight, and log hematocrit in populations of the same age

3.1. Medaka of Age 8 Months

In this section we study evidence of association of length, log weight, and log hematocrit with population of fish (experimental or breeding) for fish of age 8 months.

Figure 1 displays a scatterplot of length versus log weight for experimental population (o's) and the breeding population (+'s). Note that the breeding population has 4 fish of length 32mm whereas the maximum length for the experimental population is 31mm.

Fish of Age 8 Months

Population	Number of Fish	Mean		
		length	log weight	log hematocrit
experimental	43	27.77	5.94	3.82
breeding	52	27.75	5.89	3.81

The two sample covariance matrices are

Covariance Matrix
Experimental Population Age 8 Months (43 fish)

	length	log weight	log hematocrit
length	4.33	0.38	0.04
log weight	0.38	0.06	0.01
log hematocrit	0.04	0.01	0.02

Covariance Matrix
Breeding Population Age 8 Months (52 fish)

	length	log weight	log hematocrit
length	6.19	0.52	0.08
log weight	0.52	0.07	0.003
log hematocrit	0.08	0.003	0.03

Note that variance of the lengths in the breeding population, 6.19, is larger than that for the experimental population.

The null hypothesis that the mean vector of length, log weight, and log hematocrit are equal for the two populations cannot be rejected ($p = 0.49$).

Figure 2 presents histograms of the linear combination $a_1^*(\text{length}) + a_2^*(\text{log weight}) + a_3^*(\text{log hematocrit})$ which maximizes the discrepancy between the experimental and breeding populations. In this case

$$a_1^* = -0.18, \quad a_2^* = 2.14, \quad a_3^* = 0.13.$$

An analysis of variance for equality of mean length does not reject the null hypothesis of equal means ($p = 0.97$, $F = .007$, $df_B = 1$, $df_W = 93$). An analysis of variance for equality of mean log weights does not reject the null hypothesis of equal means ($p = 0.33$, $F = 0.95$, $df_B = 1$, $df_W = 93$). An analysis of variance for equality of log hematocrit does not reject the null hypothesis of equal means ($p = 0.85$, $F = 0.036$, $df_B = 1$, $df_W = 93$).

Conclusion. The mean vectors for length, log weight, and log hematocrit are not statistically significantly different ($p = 0.49$) for the experimental and breeding populations of medaka of age 8 months.

3.2. Medaka of Age 6 Months

In this section we consider lengths, weights, and log hematocrit for medaka of 6 months of age.

Fish of Age 6 Months

Population	Number of Fish	Mean length	Mean log weight	Mean log hematocrit
experimental	31	26.29	5.72	3.88
breeding	32	26.91	5.82	3.78

Sample Covariance Matrix
Experimental Population (6 Months)

	length	log weight	log hematocrit
length	3.28	0.28	0.008
log weight	0.28	0.03	0.003
log hematocrit	0.008	0.003	0.03

Sample Covariance Matrix
Breeding Population (6 Months)

	length	log weight	log hematocrit
length	5.83	0.62	0.26
log weight	0.62	0.08	0.03
log hematocrit	0.26	0.03	0.04

The null hypothesis that the mean vectors of the two populations are equal is rejected (p-value = 0.03). The linear combination of the measurements that results in the largest discrepancy between the experimental and breeding populations is

$$a_1^*(\text{length}) + a_2^*(\log \text{weight}) + a_3^*(\log \text{hematocrit})$$

with

$$a_1^* = 0.11, \quad a_2^* = -3.62, \quad a_3^* = 3.72.$$

An analysis of variance for equality of the mean length for the breeding and experimental populations does not reject the null hypothesis that the means are equal (p-value = 0.26 with $F = 1.31$, $df_B = 1$, $df_W = 61$). An analysis of variance for equality of the mean log weight for the two populations does not reject the null hypothesis that the means are equal (p-value = 0.12 with $F = 2.52$, $df_B = 1$, $df_W = 61$). An analysis of variance for equality of the mean log hematocrit for the populations barely rejects the null hypotheses of equal means ($p = 0.046$ with $F = 4.14$, $df_B = 1$, $df_W = 61$). The mean hematocrit level for the breeding population is smaller than that for the experimental population.

Figure 3 presents a scatterplot of log weight and log hematocrit for the two populations (o = experimental population and + = breeding population). Note the one + on the left which is away from the major point cloud. Also note the predominance of +'s in the lower portion of the plot.

Figure 4 displays histograms of the linear combination of the measurements that results in the greatest discrepancy between the experimental and breeding populations. Note that the histogram for the experimental population has a suggestion of bimodality; the bimodality casts doubt on the multivariate normal assumption for the data.

Figure 5 displays the linear combination for the experimental population by health screen. Health screen 4 has 3 of the low values and health screen 3 has 1 of the low values.

Figure 6 displays a scatterplot of the linear combination versus log hematocrit for the experimental population at age 6 months. Note the 4 points that lie away from the major point cloud.

Conclusion. There is a statistically significant ($p = 0.03$) difference in the mean vectors of length, log weight, and log hematocrit for the breeding and experimental populations for medaka of age 6 months. It remains to determine if the difference is of biological significance.

4. Comparison of log leukocrit and log hematocrit in experimental and breeding populations.

In this section we investigate the possible association of log leukocrit and log hematocrit with the population of fish (experimental or breeding) for those fish which have measured leukocrit. We do not consider the other measurements since the ages of the sacrificed fish in each health screen in the breeding and experiment populations do not match. We are assuming that leukocrit and hematocrit values do not depend on the age of the adult fish. Figure 7 displays a scatterplot of log leukocrit and log hematocrit for the experimental population (circles) and the breeding population (pluses). Note the predominance of pluses in the lower left-hand corner; this suggests that members of the breeding population have lower log leukocrit and log hematocrit values than the experimental population.

Mean

Population	log leuk.	log hemat.
experimental	-0.15	3.828
breeding	-0.47	3.817
Difference	0.32	0.011

The two sample covariance matrices are

Covariance Matrix
Experimental Population (118 fish)

	log leuk.	log hemat.
log leuk.	0.20	-0.01
log hemat.	-0.01	0.02

Covariance Matrix
Breeding Population (87 fish)

	log leuk.	log hemat.
log leuk.	0.51	-0.01
log hemat.	-0.01	0.03

The null hypothesis that the two mean vectors are equal is rejected with $p = 0.0004$. The linear compound $U = a_1^*$ (log leukocrit) + a_2^* (log hematocrit) which gives the largest value of a t-statistic to test for equal means for the two populations uses

$$a_1 = 1.00 \text{ and } a_2 = 0.83;$$

thus the maximizing linear compound is roughly an equally weighted linear combination of the two measurements. Figure 8 displays histograms of U for each population. Note that the breeding population has a smaller mean U and a greater variability.

An analysis of variance rejects the null hypothesis of equal mean log leukocrit ($p = 0.0002$, $F = 15.66$, $df_B = 1$, $df_W = 203$). An analysis of variance does not reject the null hypothesis of equal log mean hematocrit ($p = 0.63$, $F = 0.25$, $df_B = 1$, $df_W = 203$).

Conclusion: The mean vectors of log leukocrit and log hematocrit are statistically significantly different ($p = 0.0004$). Members of the breeding population tend to have lower leukocrit levels than those of the experimental population.

REFERENCES

Chatfield, C. and A.J. Collins. *Introduction to Multivariate Analysis*. Chapman and Hall, New York, 1980.

IBM Corporation. A Graphical Statistical System (AGSS).

Twerdok, Lorraine E., M.W. Curry, J.R. Beaman, and J.T. Zelikoff. "Development of routine health monitoring methods for use with an aquatic model (Oryzias Latipes) used in immunotoxicological testing." Poster presented at SOF 34th Annual Meeting, Baltimore, MD, March 7, 1995.

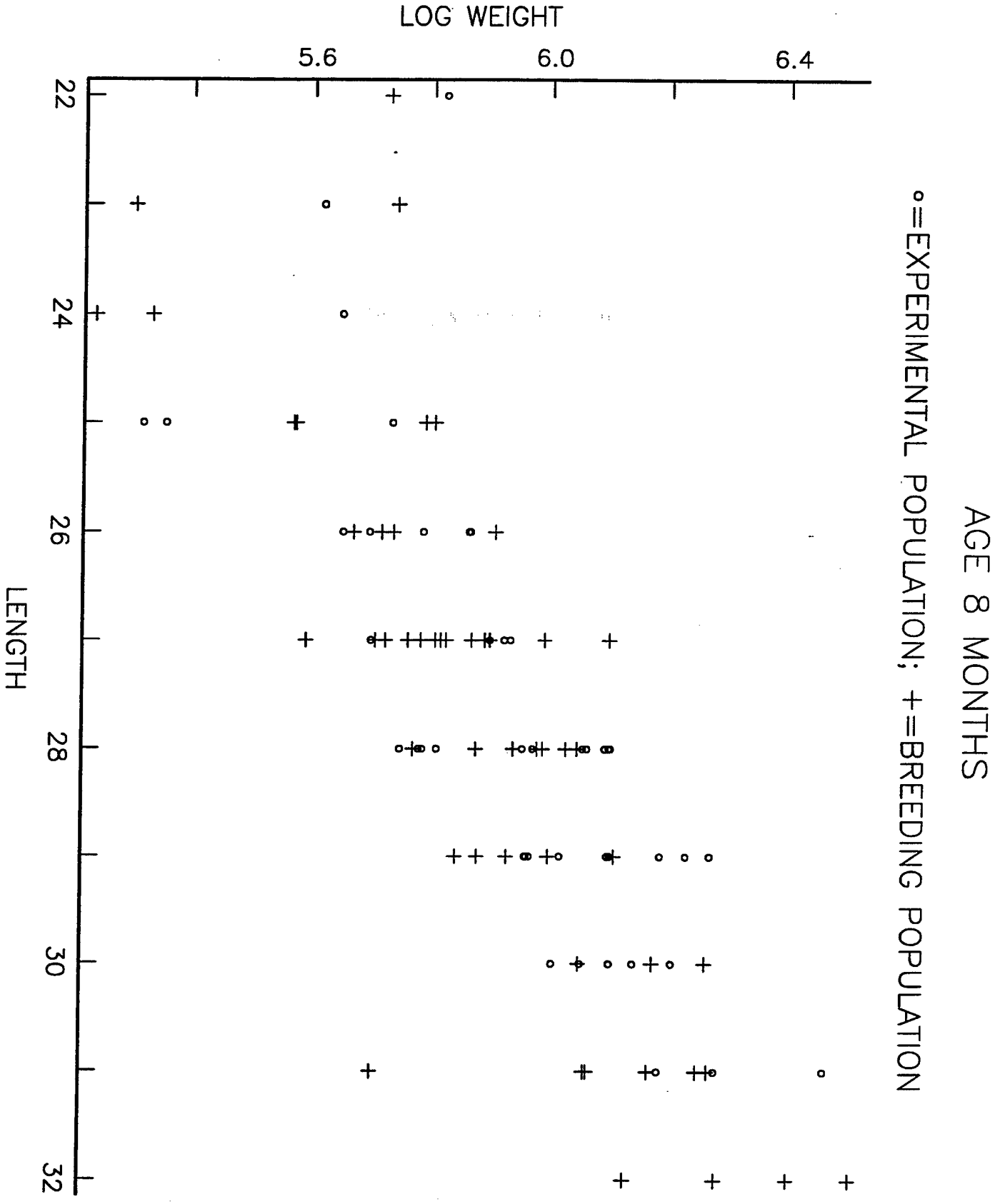

LOG WEIGHT
5.6
6.0
6.4
LENGTH
22
24
26
28
30
32
AGE 8 MONTHS
°=EXPERIMENTAL POPULATION; +=BREEDING POPULATION

Figure 1

LENGTH, LOG WEIGHT, AND LOG HEMATOCRIT FOR AGE 8 MONTHS
EXPERIMENTAL POPUL
LINEAR COMBINATION OF LENGTH, LOG WEIGHT, AND LOG HEMATOCRIT RESULTING IN THE LARGEST DISCREPANCY BETWEEN THE TWO POPUL
NO OF SAMPLES
0
5
10
15
20
5
6
7
8
9
10
LINEAR COMBINATION
BREEDING POPUL
LINEAR COMBINATION OF LENGTH, LOG WEIGHT, AND LOG HEMATOCRIT RESULTING IN THE LARGEST DISCREPANCY BETWEEN THE TWO POPUL
NO OF SAMPLES
0
5
10
15
20
5
6
7
8
9
10
LINEAR COMBINATION

Figure 2

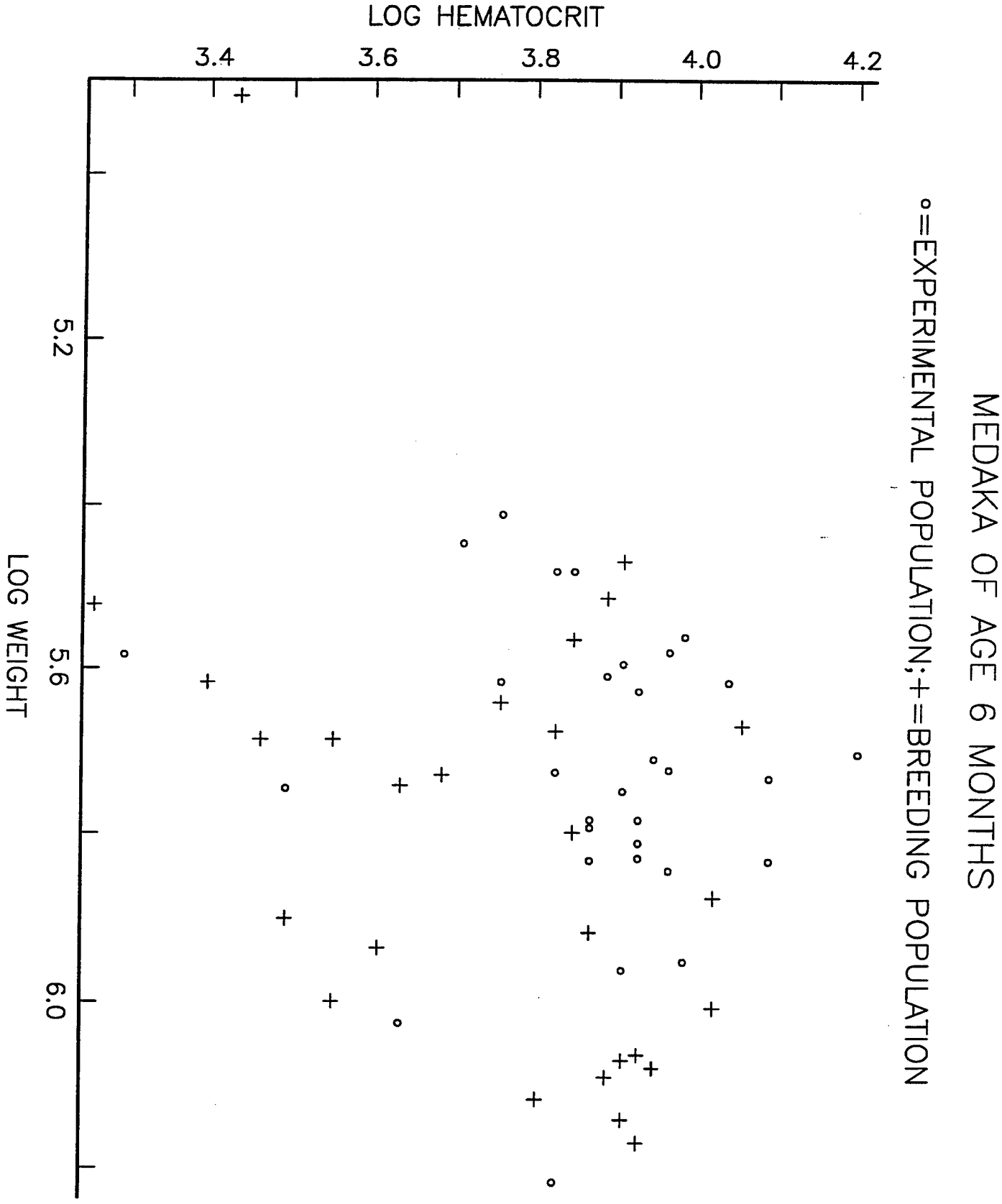

LOG HEMATOCRIT
3.4
3.6
3.8
4.0
4.2
LOG WEIGHT
5.2
5.6
6.0
MEDAKA OF AGE 6 MONTHS
°=EXPERIMENTAL POPULATION;+=BREEDING POPULATION

Figure 3

LENGTH,LOG WEIGHT, AND LOG HEMATOCRIT FOR FISH OF AGE 6 MONTHS
EXPERIMENTAL POPULATION
LINEAR COMBINATION OF LENGTH, LOG WEIGHT AND LOG HEMATOCRIT RESULTING
IN THE LARGEST DISCREPANCY BETWEEN THE TWO POPUL
NO OF SAMPLES
0
5
10
15
20
-6
-5
-4
-3
-2
-1
LINEAR COMBINATION
BREEDING POPULATION
LINEAR COMBINATION OF LENGTH, LOG WEIGHT AND LOG HEMATOCRIT RESULTING
IN THE LARGEST DISCREPANCY BETWEEN THE TWO POPUL
NO OF SAMPLES
0
5
10
15
20
-6
-5
-4
-3
-2
-1
LINEAR COMBINATION

Figure 4

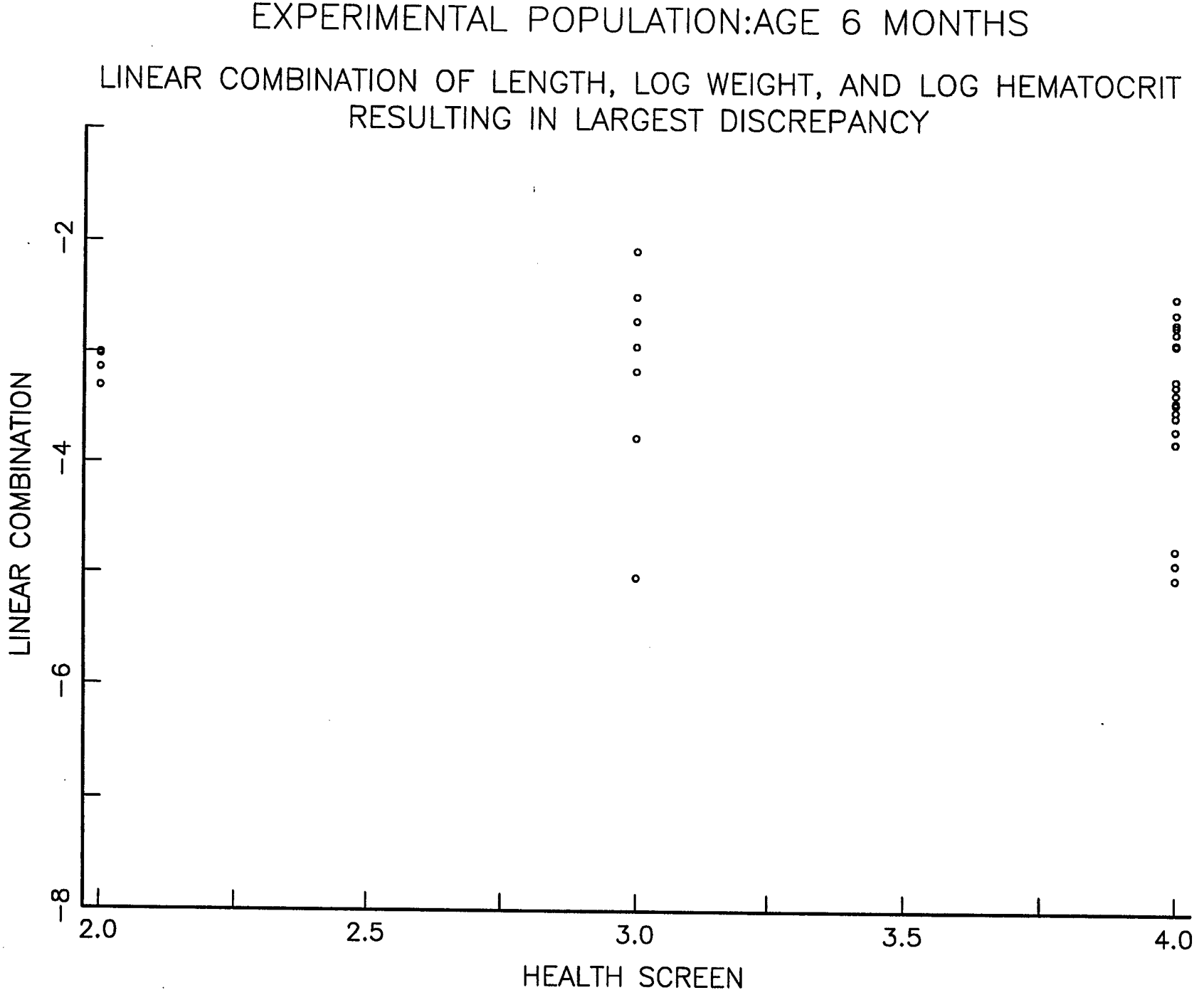

EXPERIMENTAL POPULATION:AGE 6 MONTHS
LINEAR COMBINATION OF LENGTH, LOG WEIGHT, AND LOG HEMATOCRIT
RESULTING IN LARGEST DISCREPANCY
LINEAR COMBINATION
-2
-4
-6
-8
2.0
2.5
3.0
3.5
4.0
HEALTH SCREEN

Figure 5

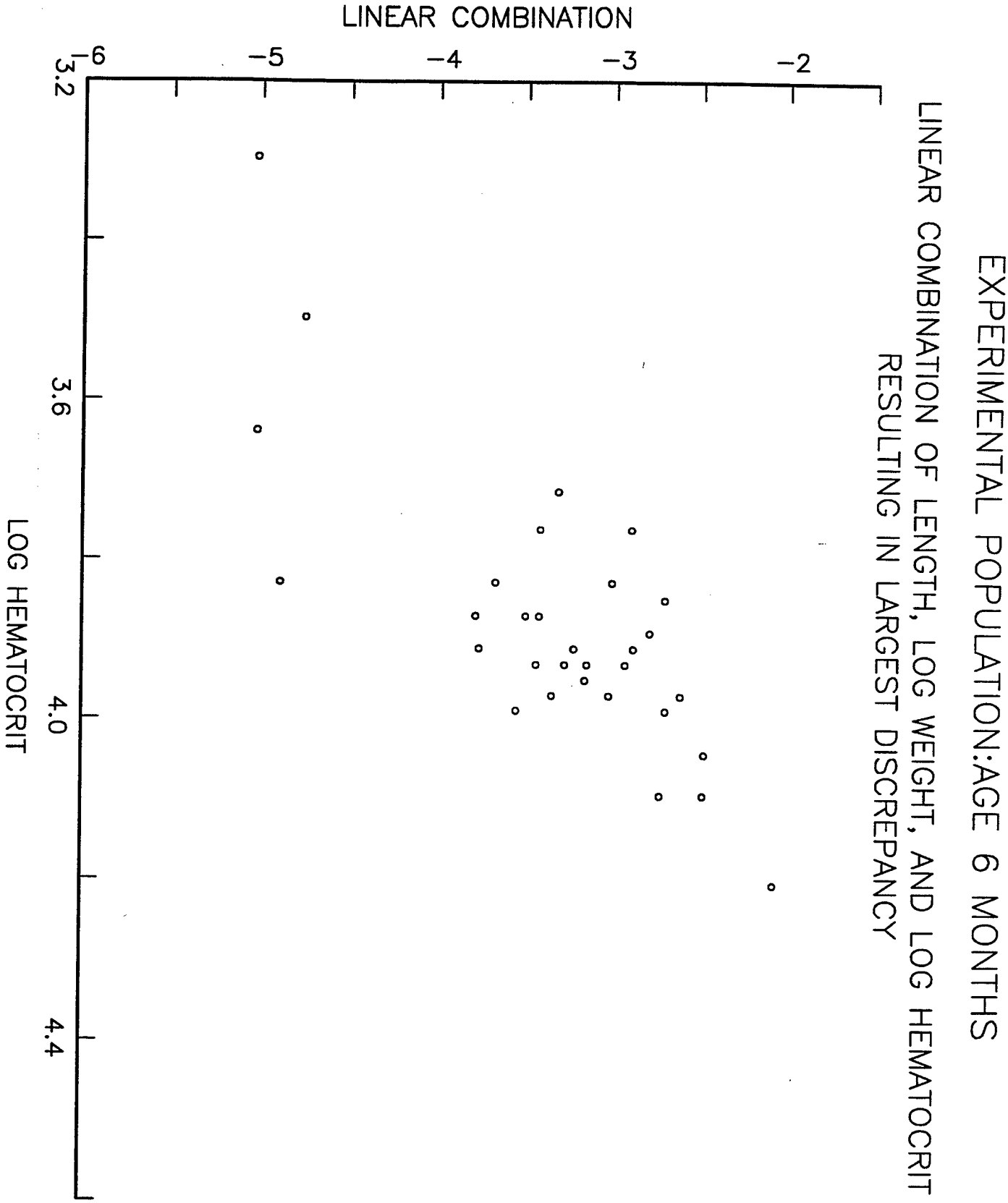

EXPERIMENTAL POPULATION:AGE 6 MONTHS
LINEAR COMBINATION OF LENGTH, LOG WEIGHT, AND LOG HEMATOCRIT
RESULTING IN LARGEST DISCREPANCY
LINEAR COMBINATION
−6
−5
−4
−3
−2
LOG HEMATOCRIT
3.2
3.6
4.0
4.4

Figure 6

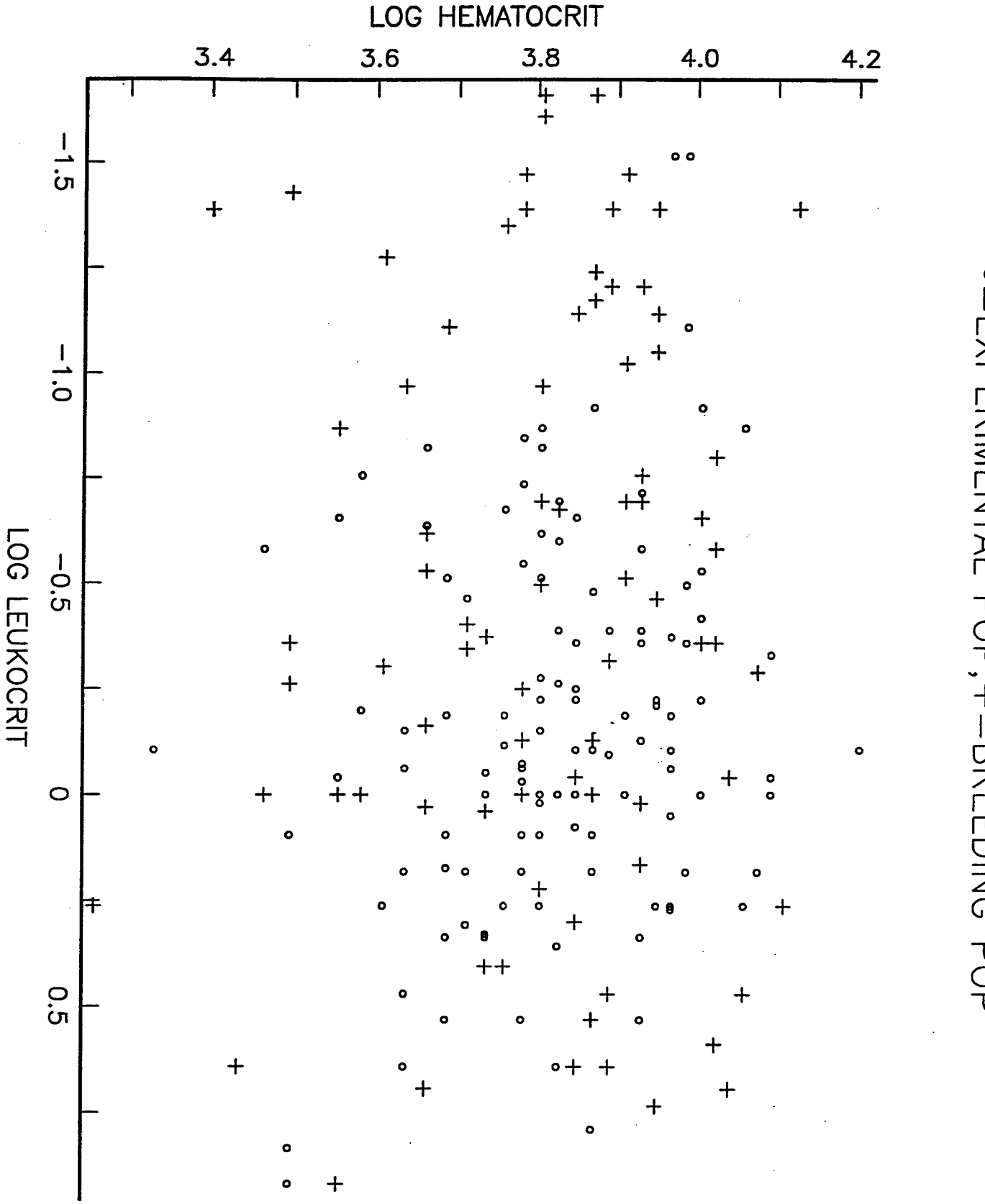

LOG HEMATOCRIT
3.4
3.6
3.8
4.0
4.2
LOG LEUKOCRIT
-1.5
-1.0
-0.5
0
0.5
°=EXPERIMENTAL POP;+=BREEDING POP

Figure 7

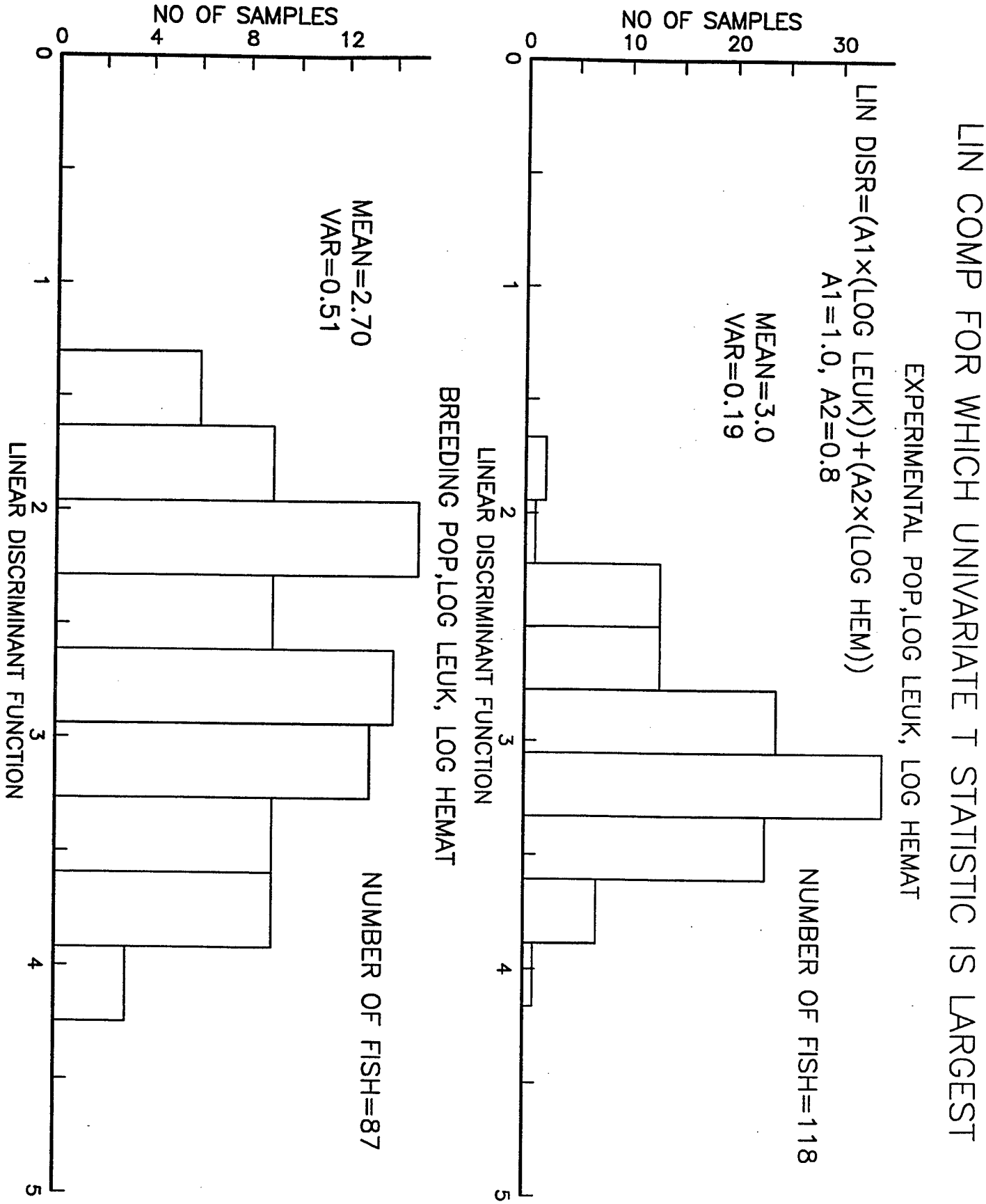

LIN COMP FOR WHICH UNIVARIATE T STATISTIC IS LARGEST
EXPERIMENTAL POP,LOG LEUK, LOG HEMAT
LIN DISR=(A1×(LOG LEUK))+(A2×(LOG HEM))
A1=1.0, A2=0.8
MEAN=3.0
VAR=0.19
NUMBER OF FISH=118
NO OF SAMPLES
0
10
20
30
LINEAR DISCRIMINANT FUNCTION
BREEDING POP,LOG LEUK, LOG HEMAT
MEAN=2.70
VAR=0.51
NUMBER OF FISH=87
NO OF SAMPLES
0
4
8
12
LINEAR DISCRIMINANT FUNCTION
0
1
2
3
4
5

Figure 8

Analysis of Some Pathology Data from the Six Month Interim Sacrifice of the West Branch Canal Creek Carcinogenicity Study with Medaka, Test 401-002R

by

D. P. Gaver and P. A. Jacobs

1. Introduction

On October 31, 1995, Margaret Toussaint, on behalf of Tom Shedd, sent us a draft copy of the pathology report of the six month interim sacrifice of the U.S. Army Biomedical Research and Development Laboratory Test 401-002R, West Branch Canal Creek Carcinogenicity Study with Medaka.

We quote from the final draft report prepared by Experimental Pathology Laboratories, Inc. (1995), hereafter referred to as EPL (1995). In the test, "groundwater was pumped from a well on-site into two flow-through diluter systems in a biomonitoring trailer. One system had water from the West Branch of Canal Creek as the dilution water. The dilution water in the second system was dechlorinated tap water. Throughout the study laboratory control medaka were maintained at Fort Detrick in well water. At 13 days of age medaka were either initiated or not initiated with 10 mg/L diethylnitrosamine (DEN) for 48 hours. Exposure to the groundwater began at 16 days of age. At six months into the study approximately 20 medaka from each exposure group were euthanized for evaluation." The study design is in Table 1.1.

TABLE 1.1

Group ID	Diluent Water	DEN (mg/L)	Groundwater (%)	No. of Fish Submitted at 6 months (Each group)
1, 2	Canal Creek	0	0	20, 20
3, 4	Canal Creek	10	0	21, 20
5, 6	Canal Creek	0	1	20, 20
7, 8	Canal Creek	10	1	20, 20
9, 10	Canal Creek	0	5	20, 21
11, 12	Canal Creek	10	5	20, 20
13, 14	Canal Creek	0	25	20, 20
15, 16	Canal Creek	10	25	20, 20
17, 18	Dechlorinated Tap	0	0	20, 20
19, 20	Dechlorinated Tap	10	0	19, 19
21, 22	Dechlorinated Tap	0	1	20, 20
23, 24	Dechlorinated Tap	10	1	20, 20
25, 26	Dechlorinated Tap	0	5	20, 20
27, 28	Dechlorinated Tap	10	5	19, 20
29, 30	Dechlorinated Tap	0	25	20, 19
31, 32	Dechlorinated Tap	10	25	20, 20
33, 34	Lab Well	0	0	20, 20
35, 36	Lab Well	10	0	19, 20

Further information concerning the study can be found in EPL (1995).

Table A.1 in Appendix A lists the number of fish from each treatment group by sex exhibiting the endpoints of Hepatocellular Adenoma (HA), Hepatocellular Carcinoma (HC), Basophilic Foci, (BF), and Eosinophilic Foci (EF).

In Section 2, logistic regression is used to study the association between the occurrence of endpoints and other covariates. The endpoints considered are the presence of hepatocellular adenoma, the presence of hepatocellular carcinoma, tthe presence of basophilic foci, and the presence of eosinophilic foci. The data appear in Table A.1 of Appendix A. The covariates considered are a constant; amount of DEN the fish is exposed to (0 mg/L or 10 mg/L); % groundwater; and indicator variables $I_{\text{Canal Creek}}$, I_{Male}, I_{Lab}; where $I_{\text{Canal Creek}} = 1$ if the diluent

water is from Canal Creek and 0 otherwise; I_{Male} equals 1 if the animal is male and 0 otherwise; I_{Lab} equals 1 if the diluent water is lab water and 0 otherwise. An association between a covariate and the presence of an endpoint is considered to be statistically significant if the parameter estimate is greater than 2 standard deviations away from 0. The results are summarized as follows.

1. The fish exposed to DEN have a statistically significant greater probability of exhibiting each endpoint than fish not exposed to DEN.

2. For animals not exposed to DEN, there is no statistical evidence that the occurrence of any of the endpoints is associated with the type of diluent water, the sex of the animal, or the % groundwater.

3. For animals exposed to DEN:
   a. there is no statistical evidence that the occurrence of hepatocellular carcinoma is associated with the type of diluent water, the sex of the animal, nor the % groundwater;
   b. the probability of an animal having hepatocellular adenoma is greater for those fish in Canal Creek diluent water than for the other diluent waters;
   c. the probability of an animal having basophilic foci is decreased if the animal is male and is decreased if the diluent is Ft. Detrick well water;
   d. the probability of an animal having eosinophilic foci is increased if the animal is male. It is also increased with an increase in % groundwater.

Some of the endpoints are categorical: 0 = not present, 1 = minimal, 2 = slight/mild, 3 = moderate, 4 = moderately severe, 5 = severe/high. Analysis of data incorporating the categorical nature of the endpoints is reported in Sections 3 – 5. The endpoints considered are the presence or absence of hepatocellular adenoma, the category of basophilic foci, the category of eosinophilic foci, the category of cystic degeneration in the liver, and the category of hyaline material in the glomeruli of the kidney.

The Kruskal-Wallis procedure is used as an exploratory procedure to look for possible associations between endpoints. The Kruskal-Wallis statistic is a nonparametric one-way analysis of variance using ranks rather than the original measurements. Those associations that were statistically significant (p-value < 0.05) were further explored using a contingency table χ^2 test for independence. The results of the contingency table analyses are summarized below.

1. For fish in Canal Creek diluent
   a. Those fish exposed to DEN tend to have higher categories of hyaline material in the glomeruli of the kidney (p-value = 0.03), higher categories of basophilic foci (p-value = 0.002), higher categories of eosinophilic foci (p-value = 0.00004), and have greater incidence of hepatocellular adenoma (p-value = 7.6×10^{-6}) than those fish not exposed to DEN.
   b. Fish that have hepatocellular adenoma tend to have higher categories of hyaline material in glomeruli of the kidney (p-value = 0.00015) and higher categories of cystic degeneration in the liver (p-value = 0.023).
   c. Males tend to have higher categories of eosinophilic foci than the females (p-value = 0.04).
   d. Females tend to have higher categories of basophilic foci than males (p-value = 0.02).

2. For fish whose diluent is tap water
   a. Fish exposed to DEN tend to have higher categories of basophilic foci than fish not exposed to DEN, (p-value = 0.002).
   b. Fish exposed to DEN tend to have higher categories of eosinophilic foci than fish not exposed to DEN (p-value = 0.0006).

3. Fish exposed to DEN and using Canal Creek water as the diluent tend to have more hepatocellular adenoma than fish exposed to DEN and using tap water as the diluent (p-value = 0.006).

4. Fish using Canal Creek water as the diluent tend to have higher categories of hyaline material in glomeruli of the kidney than fish using tap water as the diluent, (p-value = 0.00004 for fish not exposed to DEN and p-value = 9×10^{-9} for fish exposed to DEN).

2. Logistic Regression Models

Let y_i be the number of animals exhibiting a particular endpoint out of the n_i animals in tank i. Let $x_i = (x_{i1}, x_{i2}, \ldots, x_{i,p})$ represent the values of covariates for that tank, e.g. concentration of DEN, % of groundwater, etc. A probability model for y_i is the binomial distribution

$$P\{Y_i = y_i\} = \binom{n_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{n_i - y_i}, \quad y_i = 0,1,\ldots,n_i$$

where θ_i is the probability an animal in tank i displays the particular endpoint. Often θ_i is assumed to depend on covariates in the following manner.

$$\theta_i = \frac{\exp\{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}\}}{1 + \exp\{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}\}}.$$

Such a model is called a logistic regression model; cf. Collett (1991).

Table 2.1 displays results of fitting logistic regression models to the data. The covariates used are constant, $I_{\text{Canal Creek}}$, I_{Male}, I_{Lab}, and % groundwater, where $I_{\text{Canal Creek}} = 1$ if the diluent water is from Canal Creek and 0 otherwise; I_{Male} equals 1 if the animal is male and 0 otherwise; I_{Lab} equals 1 if the diluent water is lab water and 0 otherwise. Displayed are the parameter estimates and their standard errors; also displayed is the deviance of the fitted model; cf. Collett (1991). Under certain conditions, if the model is correct, the deviance is asymptotically distributed as χ^2 with $(n - p_1)$ degrees of freedom where n is the number of binomial observations and p_1 is the number of unknown parameters included in the logistic model. Hence a measure of goodness of fit of the logistic

regression is the probability the χ^2 random variable with $(n - p_1)$ degrees of freedom is greater than or equal to the deviance; this "p-value" is also displayed in Table 2.1.

Logistic regressions were fit separately using fish that had been exposed to DEN and to those that had not. The statistics package S-PLUS, Version 3.1, was used for the estimation.

Notice that some estimates have extremely large standard errors. The large standard errors are due to the fact that the fitted model has too many parameters; cf. Collett (1991).

We will say that an estimate is significantly different from 0, if its absolute value is greater than 2 times its standard error. The standard errors of those

TABLE 2.1

Model $P\{\text{endpt}\} = \exp\{x\beta\}/[1+\exp\{x\beta\}]$

where $x\beta = \beta_0 + \beta_1(I_{\text{Canal Creek}}) + \beta_2(I_{\text{Male}}) + \beta_3(I_{\text{Lab}}) + \beta_4(\%\ \text{Groundwater})$

Endpt	DEN	Estimate (Standard Error) Const β_0	I_C β_1	I_M β_2	I_L β_3	% Gw β_4	d=deviance (df = 31)	p-values $P\{\chi^2_{31} > d\}$
HA	0	-4.12 (0.80)^s	-0.61 (0.74)	1.14 (0.82)	-7.67 (40.0)	0.00067 (0.035)	20.7	0.92
	10	-3.47 (0.46)^s	1.34 (0.42)^s	0.43 (0.38)	-5.85 (9.0)	0.0317 (0.0166)	31.8	0.43
HC	0	-14.4 (177)	11.6 (177)	-11.6 (166)	-1.86 (492)	-7.72 (25.1)	1.33	1
	10	-5.79 (1.22)^s	1.30 (0.82)	2.03 (1.07)	-4.67 (8.4)	-0.0096 (0.035)	19.2	0.95
BF	0	-5.60 (1.18)^s	1.64 (1.06)	-0.17 (0.83)	-3.44 (9.22)	0.056 (0.036)	14.6	0.99
	10	-1.55 (0.31)^s	0.36 (0.34)	-1.15 (0.36)^s	-6.98 (0.06)^s	-0.009 (0.05)	42.7	0.08
EF	0	-5.70 (1.51)^s	-0.01 (1.43)	-0.07 (1.42)	-4.39 (15.2)	0.06 (0.06)	9.1	1
	10	-3.08 (0.40)^s	0.39 (0.33)	1.22 (0.36)^s	-1.40 (1.02)	0.04 (0.01)^s	37.3	0.20

s = estimate is significantly different than 0

estimates that are significantly different than 0 are marked with an s in Table 2.1 and Table 2.2.

For the animals not exposed to DEN, the only estimate that is significantly different from 0 is that associated with the constant. Thus, there is no statistical evidence that the occurrence of any of the endpoints is associated with the covariates for those animals not exposed to DEN.

For those fish exposed to DEN, the endpoints of Hepatocellular Adenoma (HA), Basophilic Foci, (BF), and Eosinophilic Foci (EF) have parameter estimates in addition to the one associated with the constant that are significantly different than 0. The only parameter estimate that is significantly different than 0 for the endpoint hepatocellular carcinoma (HC) is that associated with the constant. Thus, there is no evidence that the occurrence of hepatocellular carcinoma is associated with any of the exploratory variables.

The results for the other endpoints for those animals exposed to DEN can be summarized as follows.

1. The probability of an animal having hepatocellular adenoma is greater for those fish in Canal Creek diluent water than for the other diluent waters.
2. The probability of an animal having basophilic foci is decreased if the animal is male and is decreased if the diluent is Ft. Detrick well water.
3. The probability of an animal having eosinophilic foci is increased if the animal is male. It also increases with an increase in % groundwater.

Table 2.2 reports the result of fitting a logistic regression for each endpoint using all the fish. The covariates in the logistic regression are a constant, amount of DEN, $I_{\text{Canal Creek}}$, I_{Male}, I_{Lab}, and % Groundwater. Note that all endpoints have an estimate for the effect of DEN which is significantly different than 0. Thus, fish exposed to DEN have a significantly higher probability of exhibiting each endpoint.

TABLE 2.2

Model $P\{\text{endpt}\} = \exp\{x\beta\}/[1+\exp\{x\beta\}]$

where

$$x\beta = \beta_0 + \beta_1(\text{DEN}) + \beta_2(I_{\text{Canal Creek}}) + \beta_3(I_{\text{Male}}) + \beta_4(I_{\text{Lab}}) + \beta_4(\%\ \text{Groundwater})$$

Endpt	Estimate (Standard Error) Const β_0	DEN β_1	I_C β_2	I_M β_3	I_L β_4	% Gw β_4	d=deviance (df = 66)	p-values $P\{\chi^2_{67} > d\}$
HA	-4.75 (0.51)s	0.16 (0.04)s	0.89 (0.35)s	0.55 (0.34)	-5.58 (5.91)	0.026 (0.014)	59.1	0.71
HC	-7.49 (1.37)s	0.225 (0.106)s	1.42 (0.80)	1.33 (0.80)	-5.01 (9.7)	-0.016 (0.03)	27.6	1.00
BF	-3.88 (0.48)s	0.21 (0.04)s	0.53 (0.32)	-0.99 (0.33)s	-6.06 (5.63)	0.003 (0.02)	62.8	0.59
EF	-6.39 (0.80)s	0.34 (0.07)s	0.36 (0.32)	1.14 (0.35)s	-1.43 (1.05)	0.037 (0.014)s	47.4	0.96

s = estimate is significantly different than 0

3. Association Between Endpoints for fish in Canal Creek Diluent

Data for 5 endpoints for fish when diluent is Canal Creek water are considered. The endpoints considered are hyaline material in the glomeruli of the kidney (H), Hepatocellular Adenoma (A), Basophilic Foci, (B), Eosinophilic Foci (E), and cystic degeneration in the liver (C). The data for endpoint A are binary; 1 = present, 0 = absent. The data for H, B, E, and C are categorical; 0 = not present, 1 = minimal, 2 = slight/mild, 3 = moderate, 4 = moderately severe, 5 = severe/high.

The Kruskal-Wallis procedure is used as an exploratory procedure to look for possible associations between endpoints. The Kruskal-Wallis statistic is a nonparametric one-way analysis of variance using ranks rather than the original measurements. The null hypothesis is that the k populations have equal location parameters; cf. Gibbons (1985). The results of the procedure using all fish

exposed to Canal Creek diluent are summarized in Table 3.1. Table 3.2 presents Kruskal-Wallis procedure results for fish exposed to DEN = 10 mg/L and Canal Creek diluent. Table 3.3 presents results for fish not exposed to DEN but exposed to Canal Creek water as the diluent.

TABLE 3.1

Results of Nonparametric One-Way Analysis of Variance

All Canal Creek Diluent Fish

End-point	Groups	Kruskal-Wallis Statistic	degrees of freedom	*p*-value
H	**DEN Level**	**9.51**	**1**	**0.002**
H	% Groundwater	1.55	3	0.67
H	Sex	0.16	1	0.69
H	**Presence of Hepatocellular Adenoma**	**19.67**	**1**	**9×10^{-6}**
H	Category of Basophilic Foci	4.11	3	0.25
H	Category of Eosinophilic Foci	2.80	3	0.42
H	Category of liver cystic degeneration	5.71	4	0.22
C	DEN Level	0.91	1	0.34
C	% Groundwater	3.45	3	0.33
C	Sex	2.96	1	0.09
C	Category of Hyaline Material in Kidney Glomeruli	8.80	4	0.07
C	Presence of Hepatocellular Adenoma	3.73	1	0.053
C	**Category of Eosinophilic Foci**	**7.89**	**3**	**0.048**
C	Category of Basophilic Foci	1.41	3	0.70
B	**DEN Level**	**14.0**	**1**	**0.0002**
B	% Groundwater	2.27	3	0.52
B	**Sex**	**8.30**	**1**	**0.004**
B	Presence of Hepatocellular Adenoma	0.89	1	0.35
B	Category of Eosinophilic Foci	1.53	3	0.68
E	**DEN Level**	**23.8**	**1**	**1×10^{-6}**
E	% Groundwater	6.28	3	0.10
E	**Sex**	**6.54**	**1**	**0.01**
E	**Presence of Hepatocellular Adenoma**	**5.32**	**1**	**0.02**
E	Category of liver cystic degeneration	8.98	4	0.06
A	**DEN Level**	**20.0**	**1**	**7.8×10^{-6}**
A	% Groundwater	2.33	3	0.51
A	Sex	0.96	1	0.33

Bold lines associated with p-value less than 0.05

H = Hyaline Material in Glomeruli of the Kidney; C = Cystic Degeneration in Liver; B = Basophilic Foci; E = Eosinophilic Foci; A = Hepatocellular Adenoma

TABLE 3.2

Results of Nonparametric One-Way Analysis of Variance

Canal Creek Diluent

DEN = 10

End-point	Groups	Kruskal-Wallis Statistic	degrees of freedom	*p*-value
E	**Sex**	**8.18**	**1**	**0.004**
H	Sex	1.30	1	0.26
A	Sex	0.95	1	0.33
C	Sex	3.15	1	0.08
B	**Sex**	**8.41**	**1**	**0.004**
A	% Groundwater	3.52	3	0.32
B	% Groundwater	1.28	3	0.73
E	% Groundwater	5.86	3	0.12
H	% Groundwater	0.78	3	0.85
C	% Groundwater	7.80	3	0.0503
H	**Presence of Hepatocellular Adenoma**	**11.70**	**1**	**0.0006**
B	Presence of Hepatocellular Adenoma	0.003	1	0.96
E	Presence of Hepatocellular Adenoma	0.87	1	0.35
C	**Presence of Hepatocellular Adenoma**	**4.13**	**1**	**0.04**
B	Category of Hyaline Material in Kidney Glomeruli	2.61	4	0.63
E	Category of Hyaline Material in Kidney Glomeruli	1.39	4	0.85
C	Category of Hyaline Material in Kidney Glomeruli	5.69	4	0.22
E	Category of Basophilic Foci	3.62	3	0.31
C	Category of Basophilic Foci	2.59	3	0.46
C	Category of Eosinophilic Foci	7.72	3	0.052

Bold lines associated with p-value less than 0.05

H = Hyaline Material in Glomeruli of the Kidney; C = Cystic Degeneration in Liver; B = Basophilic Foci; E = Eosinophilic Foci; A = Hepatocellular Adenoma

TABLE 3.3

Results of Nonparametric One-Way Analysis of Variance

Canal Creek Diluent

DEN = 0

End-point	Groups	Kruskal-Wallis Statistic	degrees of freedom	*p*-value
B	Sex	0.39	1	0.53
E	Sex	0.01	1	0.91
A	Sex	0.20	1	0.66
C	Sex	0.59	1	0.44
H	Sex	0.45	1	0.50
H	% Groundwater	1.05	3	0.79
A	% Groundwater	3.72	3	0.29
B	% Groundwater	2.23	3	0.53
E	% Groundwater	5.89	3	0.12
C	% Groundwater	1.41	3	0.70
H	Presence of Hepatocellular Adenoma	0.48	1	0.49
B	Presence of Hepatocellular Adenoma	0.10	1	0.76
E	Presence of Hepatocellular Adenoma	0.03	1	0.85
C	Presence of Hepatocellular Adenoma	0.03	1	0.86
B	Category of Hyaline Material in Kidney Glomeruli	0.54	3	0.91
E	Category of Hyaline Material in Kidney Glomeruli	0.41	3	0.94
E	Category of Basophilic Foci	0.06	2	0.97
C	Category of Basophilic Foci	0.84	2	0.66
C	Category of Eosinophilic Foci	0.01	1	0.91

H = Hyaline Material in Glomeruli of the Kidney; C = Cystic Degeneration in Liver; B = Basophilic Foci; E = Eosinophilic Foci; A = Hepatocellular Adenoma

Tables 3.4 – 3.13 display data for the cases in Table 3.1 for which the Kruskal-Wallis statistic has a *p*-value less than 0.05. Evidence for possible associations is further explored using a contingency table χ^2 test for independence.

The results of the χ^2 test for independence are summarized below.

1. Those fish exposed to DEN tend to have higher catagories of hyaline material in the glomeruli of the kidney (*p*-value = 0.03), higher categories of Basophilic Foci (*p*-value = 0.002), higher categories of Eosinophilic Foci (*p*-value = 0.00004), and have greater incidence of hepatocellular adenoma (*p*-value = 10^{-6}) than those fish not exposed to DEN.

2. Fish that have hepatocellular adenoma tend to have higher categories of hyaline material in glomeruli of the kidney (p-value = 0.00015) and higher categories of cystic degeneration in the liver (p-value = 0.02). Males tend to have higher categories of Eosinophilic Foci than the females (p-value = 0.04). Females tend to have higher categories of Basophilic Foci than males (p-value = 0.02).

TABLE 3.4

Number of Fish

	Category of Hyaline Material in Glomeruli of Kidney								
	0	1	2	3	4	5	Mean	Var.	Median
DEN = 0	134	19	7	1	0	0	0.22	0.30	0
DEN = 10	111	31	14	3	2	0	0.47	0.69	0
χ^2 Test for Independence: χ^2 = 12.4 df = 5 p-value = 0.03									

TABLE 3.5

Number of Fish

	Category of Hyaline Material in Glomeruli of Kidney								
	0	1	2	3	4	5	Mean	Var.	Median
Hepatocellular Adenoma Not present	233	38	17	3	2	0	0.30	0.47	0
Hepatocellular Adenoma Present	12	12	4	1	0	0	0.79	0.67	1
χ^2 Test for Independence: χ^2 = 24.8 df = 5 p-value = 0.00015									

TABLE 3.6

Number of Fish

	Category of Cystic Degeneration in Liver								
	0	**1**	**2**	**3**	**4**	**5**	**Mean**	**Var.**	**Median**
Hepatocellular Adenoma Not present	156	83	36	14	4	0	0.73	0.90	0
Hepatocellular Adenoma Present	11	7	10	1	0	0	1.03	0.89	1
χ^2 Test for Independence: $\chi^2 = 13.0$ df = 5 p-value = 0.023									

TABLE 3.7

Number of Fish

	Category of Basophilic Foci								
	0	**1**	**2**	**3**	**4**	**5**	**Mean**	**Var.**	**Median**
DEN = 0	156	4	1	0	0	0	0.04	0.05	0
DEN = 10	137	11	12	1	0	0	0.24	0.37	0
χ^2 Test for Independence: $\chi^2 = 18.8$ df = 5 p-value = 0.002									

TABLE 3.8

Number of Fish

	Category of Eosinophilic Foci								
	0	**1**	**2**	**3**	**4**	**5**	**Mean**	**Var.**	**Median**
DEN = 0	159	2	0	0	0	0	0.01	0.01	0
DEN = 10	134	15	9	3	0	0	0.26	0.42	0
χ^2 Test for Independence: $\chi^2 = 28.1$ df = 5 p-value = 0.00004									

TABLE 3.9

Number of Fish

	Category of Eosinophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
Hepatocellular Adenoma Not present	270	13	9	1	0	0	0.12	0.19	0
Hepatocellular Adenoma Present	23	4	0	2	0	0	0.34	0.66	0
χ^2 Test for Independence: $\chi^2 = 21.9$ df = 5 p-value = 0.0005									

TABLE 3.10

Number of Fish

	Presence of Hepatocellular Adenoma	
	No	Yes
DEN = 0	158	3
DEN = 10	135	26
χ^2 Test for Independence: $\chi^2 = 20.05$ df = 1 p-value = 7.6×10^{-6}		

TABLE 3.11

Number of Fish

	Category of Eosinophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
Male	150	12	8	2	0	0	0.20	0.32	0
Female	143	5	1	1	0	0	0.07	0.12	0
χ^2 Test for Independence: $\chi^2 = 11.4$ df = 5 p-value = 0.04									

TABLE 3.12

Number of Fish

	Category of Basophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
Male	164	3	5	0	0	0	0.08	0.13	0
Female	129	12	8	1	0	0	0.21	0.31	0
χ^2 Test for Independence: $\chi^2 = 13.8$ df = 5 p-value = 0.02									

TABLE 3.13

Number of Fish

	Category of Eosinophilic Foci				
Category of liver cystic degeneration	0	1	2	3	4
0	159	7	1	0	0
1	77	6	5	2	0
2	40	2	3	1	0
3	13	2	0	0	0
4	4	0	0	0	0
χ^2 Test for Independence: $\chi^2 = 20.9$ df = 16 p-value = 0.18					

4. Association Between Endpoints for Fish Whose Diluent is Tap Water

Data for 5 endpoints for fish whose diluent is tap water are considered. The endpoints considered are hyaline material in the glomeruli of the kidney (H), Hepatocellular Adenoma (A), Basophilic Foci, (B), Eosinophilic Foci (E), and cystic degeneration in the liver (C). The data for endpoint A are binary; 1 = present, 0 = absent. The data for H, B, E, and C are categorical; 0 = not present, 1 = minimal, 2 = slight/mild, 3 = moderate, 4 = moderately severe, 5 = severe/high.

The Kruskal-Wallis procedure was used as an exploratory procedure to look for possible associations between endpoints. The results of the procedure using all fish exposed to tap water as the diluent are summarized in Table 4.1. Table 4.2

presents Kruskal-Wallis procedure results for fish exposed to DEN = 10 mg/L and tap water diluent. Table 4.3 presents results for fish not exposed to DEN but exposed to tap water as the diluent.

TABLE 4.1

Results of Nonparametric One-Way Analysis of Variance
All Tap Water Diluent Fish

End-point	Groups	Kruskal-Wallis Statistic	degrees of freedom	*p*-value
H	DEN Level	1.36	1	0.24
H	% Groundwater	0.44	3	0.93
H	Sex	0.28	1	0.60
H	Presence of Hepatocellular Adenoma	1.42	1	0.23
H	Category of Basophilic Foci	0.41	3	0.94
H	Category of Eosinophilic Foci	1.42	3	0.23
H	Category of liver cystic degeneration	4.12	4	0.39
C	DEN Level	0.45	1	0.50
C	% Groundwater	4.14	3	0.24
C	Sex	0.69	1	0.40
C	Category of Hyaline Material in Kidney Glomeruli	0.90	2	0.64
C	Presence of Hepatocellular Adenoma	0.17	1	0.68
C	Category of Eosinophilic Foci	6.18	3	0.10
C	Category of Basophilic Foci	3.12	3	0.37
B	**DEN Level**	**14.09**	**1**	**0.0002**
B	**% Groundwater**	**13.64**	**3**	**0.003**
B	Sex	2.10	1	0.15
B	Presence of Hepatocellular Adenoma	0.076	1	0.78
B	Category of Hyaline Material in Kidney Glomeruli	0.41	2	0.82
B	Category of Eosinophilic Foci	2.16	3	0.34
B	Category of liver cystic degeneration	1.66	4	0.80
E	DEN Level	**17.6**	**1**	**0.0003**
E	% Groundwater	5.51	3	0.14
E	**Sex**	**4.44**	**1**	**0.04**
E	Presence of Hepatocellular Adenoma	1.15	1	0.28
E	**Category of liver cystic degeneration**	**21.4**	**4**	**0.0003**

Bold lines associated with p-value less than 0.05

H = Hyaline Material in Glomeruli of the Kidney; C = Cystic Degeneration in Liver; B = Basophilic Foci; E = Eosinophilic Foci; A = Hepatocellular Adenoma

TABLE 4.2

Results of Nonparametric One-Way Analysis of Variance

Tap Water Diluent

DEN = 10

End-point	Groups	Kruskal-Wallis Statistic	degrees of freedom	*p*-value
E	Sex	2.40	1	0.12
H	Sex	0.70	1	0.40
A	Sex	2.11	1	0.15
C	Sex	0.55	1	0.46
B	**Sex**	**5.11**	**1**	**0.02**
A	% Groundwater	3.91	3	0.27
B	**% Groundwater**	**16.39**	**3**	**0.0009**
E	% Groundwater	4.22	3	0.24
H	% Groundwater	2.15	3	0.54
C	% Groundwater	7.55	3	0.06
H	Presence of Hepatocellular Adenoma	1.57	1	0.21
B	Presence of Hepatocellular Adenoma	0.004	1	0.95
E	Presence of Hepatocellular Adenoma	0.43	1	0.51
C	Presence of Hepatocellular Adenoma	0.27	1	0.60
B	Category of Hyaline Material in Kidney Glomeruli	0.58	2	0.75
E	Category of Hyaline Material in Kidney Glomeruli	0.62	2	0.73
C	Category of Hyaline Material in Kidney Glomeruli	1.01	2	0.60
E	Category of Basophilic Foci	1.15	3	0.76
C	Category of Basophilic Foci	2.33	3	0.51
C	Category of Eosinophilic Foci	5.18	3	0.16

Bold lines associated with p-value less than 0.05

H = Hyaline Material in Glomeruli of the Kidney; C = Cystic Degeneration in Liver; B = Basophilic Foci; E = Eosinophilic Foci; A = Hepatocellular Adenoma

TABLE 4.3

Results of Nonparametric One-Way Analysis of Variance

Tap Water Diluent

DEN = 0

End-point	Groups	Kruskal-Wallis Statistic	degrees of freedom	*p*-value
B	Sex	1.12	1	0.29
E	Sex	1.12	1	0.29
A	Sex	2.22	1	0.14
C	Sex	0.34	1	0.56
H	Sex	0.006	1	0.94
H	% Groundwater	6.0	3	0.11
A	% Groundwater	0.66	3	0.88
B	% Groundwater	3.08	3	0.38
E	% Groundwater	3.08	3	0.38
C	% Groundwater	0.36	3	0.95
H	Presence of Hepatocellular Adenoma	0.07	1	0.80
B	Presence of Hepatocellular Adenoma	0.03	1	0.86
E	Presence of Hepatocellular Adenoma	0.03	1	0.86
C	Presence of Hepatocellular Adenoma	0.009	1	0.93
B	Category of Hyaline Material in Kidney Glomeruli	0.013	1	0.91
E	Category of Hyaline Material in Kidney Glomeruli	0.013	1	0.91
C	Category of Hyaline Material in Kidney Glomeruli	0.003	1	0.96
E	Category of Basophilic Foci	0.006	1	0.94
C	Category of Basophilic Foci	0.64	1	0.42
C	Category of Eosinophilic Foci	2.62	1	0.11

H = Hyaline Material in Glomeruli of the Kidney; C = Cystic Degeneration in Liver; B = Basophilic Foci; E = Eosinophilic Foci; A = Hepatocellular Adenoma

Tables 4.4 – 4.8 display data for those cases in Table 4.1 whose *p*-value is less than 0.05. A contingency table χ^2 test for independence is done to explore possible associations.

TABLE 4.4

Number of Fish

	Category of Basophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
DEN = 0	158	0	1	0	0	0	0.13	0.03	0
DEN = 10	141	8	4	4	0	0	0.18	0.35	0
χ^2 Test for Independence: $\chi^2 = 18.8$ df = 5 *p*-value = 0.002									

TABLE 4.5

Number of Fish

	Category of Basophilic Foci					
% Groundwater	0	1	2	3	4	5
0	68	4	2	4	0	0
1	80	0	0	0	0	0
5	76	2	1	0	0	0
25	75	2	2	0	0	0
	χ^2 Test for Independence: $\chi^2 = 27.3$ df = 15 p-value = 0.03					
without 1% Gw:	χ^2 Test for Independence: $\chi^2 = 16.0$ df = 10 p-value = 0.10					

TABLE 4.6

Number of Fish

	Category of Eosinophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
DEN = 0	158	1	0	0	0	0	0	0	0
DEN = 10	138	9	4	6	0	0	0.22	0.46	0
χ^2 Test for Independence: $\chi^2 = 21.7$ df = 5 p-value = 0.0006									

TABLE 4.7

Number of Fish

	Category of Eosinophilic Foci							
Category of liver cystic degeneration	0	1	2	3	4	Mean	Var.	Median
0	162	4	1	3	0	0.09	0.20	0
1	91	4	0	1	0	0.07	0.13	0
2	35	1	1	2	0	0.23	0.55	0
3	7	1	0	0	0	0.13	0.13	0
4	1	0	2	0	0	1.33	1.33	2
	χ^2 Test for Independence: $\chi^2 = 115.9$ df = 16 p-value = 0							
without category 4 of liver cystic degeneration:	χ^2 Test for Independence: $\chi^2 = 8.61$ df = 9 p-value = 0.47							

TABLE 4.8

Number of Fish

	Category of Eosinophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
Male	151	7	2	6	0	0	0.17	0.39	0
Female	145	3	2	0	0	0	0.05	0.07	0
χ^2 Test for Independence: $\chi^2 = 10.93$ df = 5 p-value = 0.05									

The results of the χ^2 test for independence can be summarized as follows. The exposure to DEN statistically significantly increases the category of Basophilic (p-value = 0.002) and Eosinophilic foci (p-value = 0.0006) for fish in tap water diluent.

5. Endpoint Comparison for Fish in Canal Creek Diluent and Tap Water Diluent

In this section we report comparisons of categories of endpoints for fish from Canal Creek diluent and tap water diluent. Tables 5.1 – 5.10 present data for the numbers of fish in each endpoint category versus diluent for those fish exposed to DEN and those fish not exposed to DEN. The χ^2 test for independence is again invoked.

TABLE 5.1

Number of Fish

DEN = 0

	Category of Basophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
Tap Water Diluent	158	0	1	0	0	0	0.01	0.03	0
Canal Creek Diluent	156	4	1	0	0	0	0.04	0.05	0
χ^2 Test for Independence: $\chi^2 = 10.0$ df = 5 p-value = 0.08									

TABLE 5.2
Number of Fish
DEN = 10

	Category of Basophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
Tap Water Diluent	141	8	4	4	0	0	0.18	0.35	0
Canal Creek Diluent	137	11	12	1	0	0	0.24	0.37	0
χ^2 Test for Independence: $\chi^2 = 10.3$ df = 5 p-value = 0.07									

TABLE 5.3
Number of Fish
DEN = 0

	Category of Eosinophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
Tap Water Diluent	158	1	0	0	0	0	0.006	0.006	0
Canal Creek Diluent	159	2	0	0	0	0	0.01	0.01	0
χ^2 Test for Independence: $\chi^2 = 8.3$ df = 5 p-value = 0.14									

TABLE 5.4
Number of Fish
DEN = 10

	Category of Eosinophilic Foci								
	0	1	2	3	4	5	Mean	Var.	Median
Tap Water Diluent	138	9	4	6	0	0	0.22	0.46	0
Canal Creek Diluent	134	15	9	3	0	0	0.26	0.42	0
χ^2 Test for Independence: $\chi^2 = 8.43$ df = 5 p-value = 0.13									

TABLE 5.5
Number of Fish
DEN = 0

	Category of Cystic Degeneration								
	0	1	2	3	4	5	Mean	Var.	Median
Tap Water Diluent	87	51	16	4	1	0	0.62	0.67	0
Canal Creek Diluent	91	35	24	7	4	0	0.74	1.1	0
χ^2 Test for Independence: $\chi^2 = 9.3$ df = 5 p-value = 0.10									

TABLE 5.6
Number of Fish
DEN = 10

	Category of Cystic Degeneration								
	0	**1**	**2**	**3**	**4**	**5**	**Mean**	**Var.**	**Median**
Tap Water Diluent	83	45	23	4	2	0	0.71	0.81	0
Canal Creek Diluent	76	55	22	8	0	0	0.76	0.76	1
χ^2 Test for Independence: $\chi^2 = 6.6$ df = 5 p-value = 0.25									

TABLE 5.7
Number of Fish
DEN = 0

	Presence of Hepatocellular Adenoma	
	Not Present	**Present**
Tap Water Diluent	154	5
Canal Creek Diluent	158	3
χ^2 Test for Independence: $\chi^2 = 0.54$ df = 1 p-value = 0.46		

TABLE 5.8
Number of Fish
DEN = 10

	Presence of Hepatocellular Adenoma	
	Not Present	**Present**
Tap Water Diluent	147	10
Canal Creek Diluent	135	26
χ^2 Test for Independence: $\chi^2 = 7.6$ df = 1 p-value = 0.006		

TABLE 5.9
Number of Fish
DEN = 0

	Category of Hyaline Material in Kidney Glomeruli							
	0	**1**	**2**	**3**	**4**	**5**	**Mean**	**Var.**
Tap Water Diluent	157	2	0	0	0	0	0.01	0.01
Canal Creek Diluent	134	19	7	1	0	0	0.22	0.30
χ^2 Test for Independence: $\chi^2 = 27.6$ df = 5 p-value = 0.00004								

TABLE 5.10
Number of Fish
DEN = 10

	Category of Hyaline Material in Kidney Glomeruli							
	0	**1**	**2**	**3**	**4**	**5**	**Mean**	**Var.**
Tap Water Diluent	154	4	1	0	0	0	0.04	0.05
Canal Creek Diluent	111	31	14	3	2	0	0.47	0.69
χ^2 Test for Independence: $\chi^2 = 46.1$ df = 5 p-value = 9×10^{-9}								

The results of the χ^2 test for independence are as follows. There is evidence that fish in Canal Creek diluent tend to have higher categories of hyaline material in glomeruli of the kidney than fish in tap water diluent (p-value = 0.00004 for fish not exposed to DEN and p-value = 10^{-9} for fish exposed to DEN). Fish in Canal Creek diluent that have been exposed to DEN have a greater chance of having hepatocellular adenoma than fish in tap water diluent that have been exposed to DEN (p-value = 0.006).

References

Collett, D. *Modelling Binary Data*. Chapman & Hall, New York, NY, 1991.

Experimental Pathology Laboratories, Inc. "U.S. Army Biomedical Research and Development Laboratory Test 401-002R, EPL Project Number 406-035, West Branch Canal Creek Carcinogenicity Study with Medaka, six month interim sacrifice, pathology report, 3 volumes," Final Draft, September 21, 1995.

Gibbons, J.D. *Nonparametric Methods for Quantitative Analysis*, Second Edition, American Sciences Press, Inc., Columbus, OH, 1985.

Statistical Sciences, Inc. S-PLUS for Windows, Version 3.1, Statistical Sciences, Seattle, WA, 1993.

APPENDIX A

TABLE A.1

Number of Fish with
Hepatocellular Adenoma (HA), Hepatocellular Carcinoma (HC),
Basophilic Foci, (BF), and Eosinophilic Foci (EF)
(* denotes 1 fish died prematurely)
Diluent Water: C = Canal Creek; T = Dechlorinated Tap; L = Lab Well

Group ID	Diluent Water	DEN (mg/L)	Groundwater (%)	No. Males	No. Females	No. with HA	No. with HC	No. with BF	No. with EF
1	C	0	0	12	—	0	0	0	0
				—	8	0	0	0	0
2	C	0	0	11	—	0	0	0	0
				—	9	0	1	0	0
3	C	10	0	10	—	3	1	0	2
				—	11	1	0	4	0
4	C	10	0	10	—	1	2	0	4
				—	10	2	0	0	1
5	C	0	1	12	—	1	0	0	0
				—	8	0	0	0	0
6	C	0	1	7	—	1	0	0	0
				—	13	0	0	1	0
7	C	10	1	11	—	2	0	1	1
				—	9	0	0	1	0
8	C	10	1	8	—	1	0	1	0
				—	12	3	0	4	1
9	C	0	5	15*	—	0	0	0	0
				—	5	0	0	0	0
10	C	0	5	7	—	0	0	1	0
				—	14	1	0	1	1
11	C	10	5	12	—	0	0	1	5
				—	8	0	0	2	1
12	C	10	5	9	—	2	1	0	2
				—	11	2	1	4	0
13	C	0	25	11	—	0	0	1	0
				—	9	0	0	0	0
14	C	0	25	12	—	0	0	0	0
				—	8	0	0	1	0
15	C	10	25	14	—	2	2	1	4
				—	6	1	0	1	1
16	C	10	25	11	—	6	0	2	3
				—	9	1	0	2	2
17	T	0	0	7	—	1	0	0	0
				—	13	0	0	0	0
18	T	0	0	12	—	0	0	0	0
				—	8	0	0	0	0

TABLE A.1 (Continued)

Group ID	Diluent Water	DEN (mg/L)	Groundwater (%)	No. Males	No. Females	No. with HA	No. with HC	No. with BF	No. with EF
19	T	10	0	10	—	0	0	1	2
				—	9	1	0	3	0
20	T	10	0	11	—	0	0	3	0
				—	8	0	0	4	0
21	T	0	1	8	—	0	0	0	0
				—	12	0	0	0	0
22	T	0	1	10	—	1	0	0	0
				—	10	0	0	0	0
23	T	10	1	11	—	1	1	0	4
				—	9	0	0	0	1
24	T	10	1	13	—	0	0	0	0
				—	7	1	0	0	0
25	T	0	5	7	—	0	0	0	0
				—	13	0	0	0	0
26	T	0	5	12	—	1	0	0	0
				—	8	0	0	0	0
27	T	10	5	10*	—	0	0	1	0
				—	9	0	0	1	0
28	T	10	5	10	—	1	1	0	3
				—	10	0	0	1	1
29	T	0	25	6	—	0	0	0	0
				—	14	0	0	0	0
30	T	0	25	13	—	1	0	1	1
				—	6*	1	0	0	0
31	T	10	25	10	—	2	0	1	3
				—	10	0	0	2	1
32	T	10	25	9	—	1	0	0	2
				—	11	1	0	0	2
33	L	0	0	10	—	0	0	0	0
				—	10	0	0	0	0
34	L	0	0	10	—	0	0	0	0
				—	10	0	0	0	0
35	L	10	0	9	—	0	0	0	0
				—	10	0	0	0	0
36	L	10	0	13	—	0	0	0	1
				—	7	0	0	0	0

DISTRIBUTION LIST

1. Research Office (Code 09) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Dudley Knox Library (Code 013) 2
   Naval Postgraduate School
   Monterey, CA 93943-5002

3. Defense Technical Information Center 2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

4. Therese Bilodeau, Editorial Assistant 1
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Prof. Donald P. Gaver (Code OR/Gv) 5
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Prof. Patricia A. Jacobs (Code OR/Jc) 5
   Naval Postgraduate School
   Monterey, CA 93943-5000

7. Dr. Frederic Bois 1
   IEP-MS90-3058
   Lawrence Berkeley National Laboratory
   Berkeley, CA 94720

8. Dr. James G. Burkhart 1
   MDC4-07
   Environmental Toxicology Program
   NIEHS
   Research Triangle Park, NC 27709

9. Dr. Robert Carpenter 1
   NAMRI/Navy Toxicology Detachment
   Bldg. 433, Area B
   Wright-Patterson AFB, OH 45433-6503

10. Mr. Harvey Clewell, III 1
    ICF Kaiser Engineer
    ICF Kaiser International Inc.
    1201 Gaines Ave.
    Ruston, LA 71270-3107

11. Dr. Kenny S. Crump 1
Vice President, Environment Energy Group
ICF Kaiser Engineer
ICF Kaiser International Inc.
1201 Gaines Ave.
Ruston, LA 71270-3107

12. Mr. Henry S. Gardner 1
U.S. Army Biological Research & Development Laboratory
Ft. Detrick, MD 21702-5010

13. Dr. Robert C. Jackson 1
Vice President, Research and Development
Agouron Pharmaceuticals, Inc.
3565 General Atomics Court
San Diego, CA 92121-1121

14. LCDR Paul Knechtges 1
Program Manager, Fleet Occupational Health
Naval Medical Research & Development Command
Bethesda, MD 20889-5606

15. Dr. Ralph Kodell 1
Chief, Biometry Branch
Biometry and Risk Assessment
National Center for Toxicology Research
3900 NCTR Drive
Jefferson, AR 72079

16. CAPT (Sel) Mike Lilienthal 1
Aviation Medicine/Human Performance Research
Naval Medical Research & Development Command
Bethesda, MD 20889-5044

17. Dr. Jim Petty 1
National Biological Survey
4200 New Haven Road
Columbia, MO 65201

18. Dr. Raymond S.H. Yang 1
Colorado State University
College of Veterinary Medicine and Biomedical Sciences
Dept. of Environmental Health
Fort Collins, CO 80523